23364

National Library of Canada

Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE

THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* John Angus MacDonald

TITLE OF THESIS/*TITRE DE LA THÈSE* A Study of the Relationship Between Student Achievement of Process Skills and the Mode of Instruction in Junior High School Science

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Doctor of Philosophy

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE DEGRÉ* 1974

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dr. Marshall A. Nay

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* October 1, 1974 SIGNED/*SIGNÉ* JA MacDonald

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* 5639 Rannock Ave.
Winnipeg, Manitoba R3R 0N3

NL-91 (11-73)

THE UNIVERSITY OF ALBERTA

A STUDY OF THE RELATIONSHIP BETWEEN STUDENT ACHIEVEMENT OF PROCESS SKILLS AND THE MODE OF INSTRUCTION IN JUNIOR HIGH SCHOOL SCIENCE

by

©

JOHN MACDONALD

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND

RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

FALL, 1974

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "A Study of the Relationship Between Student Achievement of Process Skills and the Mode of Instruction in Junior High School Science" submitted by John MacDonald in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Secondary Education.

M. A. Nay
Supervisor

Frank L. Weizhman

H Kass

T.O. Maguire

External Examiner

Date Sept. 20. 1974

ABSTRACT

The purposes of this study were to (a) develop evaluative procedures for a process approach junior high school science curriculum and (b) to use these procedures to obtain information about the techniques and methodologies which tended to maximize process approach objectives.

Using a theoretical process approach instructional model, an evaluation model was developed and used as the basis for the development and utilization of the evaluative procedures. In essence, the evaluation model had three main thrusts: (1) the determination of student achievement of process approach objectives, (2) the determination of the nature of the instructional input, and (3) the comparisons of achievement to instructional input.

The sample consisted of three hundred and seventy-three grade seven students in sixteen classes under the jurisdiction of six teachers. This sample was divided into six treatment groups defined according to the teacher from whom they received instruction.

Achievement was determined using tests selected, prepared or adapted for the purpose. The Test On Understanding Science Form Ew was used to measure knowledge

about science and scientists; the *Inquiry Efficiency Test* (an adaptation of the *TAB Science Test*), and two prepared tests (the *Science Reasoning Test* and the *Process of Science Test*) were used to measure student performance competencies in the process dimension.

The nature of the instructional input was determined from reports submitted by participating teachers as supplemented and verified by an observational instrument designed for the purpose.

The sample was pretested using the *COOP Science Test* and the *TOUS-Ew*. After a period of sixteen weeks the the following tests were administered: The *TOUS-Ew*, the the *Science Reasoning Test*, the *Process of Science Test*, the *Inquiry Efficiency Test*, and an attitude test known as the *How I Feel About My School Test (HIFAMS)*.

Comparisons were made among groups first using one-way analysis of variance and then using analysis of covariance in a factorial design with attitude and sex treated as categorical variables.

It was determined that after allowing for the effects of intelligence and previous science knowledge there were significant differences among groups in student performance competencies in the process dimension. There were also wide variations in presentation, implementation, and interpretation of the overall curriculum.

It was concluded that the kind of instructional input which tends to maximize process approach objectives is characterized by the following: (a) non-verbal instructional modes form a dominant part of the learning environment, (b) the teacher's questions are mainly conceptual or thought-provoking as opposed to questions of a factual nature, (c) although independent work is encouraged, structure and guidance are provided in terms of student needs, and (d) the processes of science are utilized as dictated by the subject matter under consideration.

ACKNOWLEDGEMENTS

The writer expresses appreciation to the following:

Committee members: Professor M.A. Nay (Chairman), H. Kass, T. Maguire, F. Weichman, and D. Wilson. Special appreciation is expressed to Dr. H. Kass for being readily available to provide assistance and advice.

The Edmonton Public School Board for granting permission to undertake this study.

Messrs. Al Merrick, Lloyd Pearce, Reg Roberts, Wayne Phare, and Ainsley Sankarelli for their assistance and cooperation.

David Powley, a colleague in this study, for his assistance, cooperation, and helpful comments.

My wife, Isabelle, for her encouragement, patience, and understanding.

TABLE OF CONTENTS

CHAPTER PAGE

I. THE PROBLEM: ITS BACKGROUND, NATURE AND SIGNIFICANCE 1

Introduction 1

Background of the Study 5

The Problem 10

Definitions 13

Hypotheses 15

Delimitations 18

Limitations 18

Significance of the Study 20

II. REVIEW OF RELATED LITERATURE 23

Considerations Related to the Rationale of the Experimental Curriculum 23

Introduction 23

Developmental Theory and the Importance of Structure 26

Scientific Inquiry and the Process Approach in Science Teaching 32

The Edmonton Junior High School Process-Approach Science Project 41

CHAPTER PAGE

Evaluation in Science Education With Emphasis on the Process Dimension . . . 53
Early Attempts at Measuring Process Competencies 53
Tests Which Measure Knowledge About Science and Scientists 58
Tests Which Measure the Ability to Use the Processes of Science 64
Summary 75
Techniques in Classroom Observations . . . 77
III. THE DESIGN AND DEVELOPMENT OF EVALUATIVE PROCEDURES 91
The Evaluation Model 91
The Sample 99
Testing Procedures and Treatments 104
Tests and Instruments 107
Validation Procedures 122
Summary 154
The Design of the Study 156
IV. RESULTS 160
Preliminary Data 160
Comparisons Among Treatment Groups 166

CHAPTER PAGE

Results of Observational Procedures and Allied Analyses 201
Summary 224
V. SUMMARY AND GENERALIZATIONS 226
Results 227
Discussion of Results 230
For Further Research 243
BIBLIOGRAPHY 247
APPENDIX A. An Inventory of Processes in Scientific Inquiry 262
APPENDIX B. Teacher Report Formats and Instructions to Teachers 268
APPENDIX C. Tests and Instruments 272
APPENDIX D. Sample Pages From the Experimental Curriculum 333
APPENDIX E. The Attributes of a Good Teacher 340

LIST OF TABLES

TABLE PAGE

I. Distribution of Students 103

II. Pilot Study Data for Process of Science Test and Science Reasoning Test 116

III. Chi-square Calculations to Test Observer Agreement 121

IV. Correlation of Science Test Scores and Teachers' Ratings 128

V. A Summary of the Composite Relationship Between Science Test Scores and Teachers' Ratings 131

VI. Prediction of Teachers' Rating Scores Using SRT1, SRT2, SRT3, PST1, PST2, and IET Scores as Variables 133

VII. Prediction of Teachers' Rating Scores Using COOP Test, SRT, PST, and IET Test Scores as Variables 134

VIII. Factor Solution of HIFAMS with Varimox Rotation 137

IX. Correlation of HIFAMS Test Result and Teachers' Ratings 144

TABLE PAGE

X. A Summary of the Composite Relationships Between HIFAMS Test Scores and Teachers' Ratings 144
XI. Alpha Factor Solution for Twenty Variables (Varimax Rotation) 150
XII. Comparisons Among the Means of the Treatment Groups on the I.Q. (Nonverbal) Test Scores 162
XIII. Comparisons Among the Means of the Treatment Groups on the I.Q. (Verbal) Test Scores 163
XIV. Comparisons Among the Means of the Treatment Groups on the COOP Science Test Scores 164
XV. Newman-Keuls Comparison Between Ordered Means for the COOP Science Test 165
XVI. Comparisons Among the Means of the Treatment Groups on the HIFAMS(1) Scores 168
XVII. Comparisons Among the Means of the Treatment Groups on the HIFAMS(2) Scores 169
XVIII. Comparisons Among the Means of the Treatment Groups on the HIFAMS(3) Scores 170

TABLE PAGE

XIX. Comparisons Among the Means of the Treatment Groups of the HIFAMS(4) Scores 171

XX. Comparisons Among the Means of the Treatment Groups on the Process of Science Test Scores 172

XXI. Comparisons Among the Means of the Treatment Groups on the Inquiry Efficiency Test Scores 173

XXII. Comparisons Among the Means of the Treatment Groups on the Science Reasoning Test Scores 174

XXIII. Comparisons Among the Means of the Treatment Groups on TOUS Test Scores 175

XXIV. Comparisons of Treatment Groups for Differences Among Adjusted Means on The Process of Science Test 181

XXV. Comparisons of Treatment Groups for Differences Among Adjusted Means on The Inquiry Efficiency Test 182

XXVI. Comparisons Among Treatment Groups for Differences Among Adjusted Means on The Science Reasoning Test 183

TABLE PAGE

XXVII. Comparisons Among Treatment Groups for Differences Among Adjusted Means on The TOUS Test (Posttest) 184

XXVIII. A Comparison of Adjusted Treatment Means on the Science Tests 187

XXIX. Significant Differences in Adjusted Group Means on the Process of Science Test and the Inquiry Efficiency Test 189

XXX. Normal Deviates for TOUS Test Items Estimating Change Between Pretest and Posttest 192

XXXI. Normal Deviates for Test Items Estimating Change Between Pretest and Posttest for the Complete Sample 194

XXXII. A Comparison of Teachers in Terms of Time Devoted to Different Instructional Modes 202

XXXIII. A Comparison of Teachers in Terms of Percentage of Total Processes Spent in Each Gross Process and in Each Major Process 203

TABLE PAGE

XXXIV. Percentage of Time Devoted to Each Process During Formal Observation 208

XXXV. Summary of Teacher Perception of Processes Taught Compared to the Criterion . 213

XXXVI. Percentages and Ranking of Time Devoted to Each Process During Long Term Period Contrasted to Percentages of Time During Formal Observation 215

XXXVII. A Comparison of the Ranking of the Groups on Selected Instruments and Procedures . . 221

LIST OF FIGURES

FIGURE PAGE

1. An Instructional Model 94
2. The Evaluation Model 98
3. Teachers' Guide for Assigning Student Grades 124
4. A Profile of Science Achievement 177
5. Ranking of Treatments in Terms of Consistency with the Inventory 217

CHAPTER I

THE PROBLEM: ITS BACKGROUND, NATURE, AND SIGNIFICANCE

I. INTRODUCTION

The reform movements in science education which produced science curricula such as PSSC (Physical Science Study Committee), CHEM Study (Chemical Education Material Study), and BSCS (Biological Science Curriculum Study) now have their counterparts at the junior high school and elementary school level. Some of these new materials are based upon the idea that science includes two main dimensions; one dealing with subject matter and the other dealing with the methods of inquiry which lead to the acquisition of the subject matter.

These two dimensions are what Joseph Schwab (104) calls the substantive and syntactical domains of science and are the product and the process of science respectively. They receive varying degrees of relative emphasis in current curriculum materials depending on the philosophy of the designers of the materials. The influence of specialists within a discipline is reflected in both the suggested modes of inquiry of these materials and the selection of viewpoints around which to organize the content. Witness, for example, the BSCS materials which have been divided into blue, green and yellow versions emphasizing molecular, cellular and ecological viewpoints respectively. Although the molecular biology would seem to be the most abstract and theoretical, all versions may easily be fitted under the same broad conceptual schemes.

It would seem that if the process of science is to be emphasized within a science curriculum so that science can be interpreted within the meanings associated with the discipline, it is unrealistic to expect to accomplish this by having the student truly imitate the scientist. The variety of disciplines and the fragmentation within each discipline makes this impossible. It is assumed, however, that it is possible for the student to carry out some of the activities which are common to all scientists. A comprehensive list of these activities is given in the Inventory of Processes of Scientific Inquiry developed by the Edmonton Junior High School Project under the direction of M.A. Nay (80).

The belief that students learn science better when operating within a framework similar to that of the practising scientist has not been proven but it complements considerations from psychological studies that learning is best retained if it follows a pattern (16). Such a pattern depends upon the structure of the subject matter and is unique to the learner's way of structuring the subject. This, and the belief that involvement is crucial to the learning process leads to the dual problem of how to best structure the subject matter to fit the individual student and how to bring about true student involvement.

Progress towards a solution of these problems has not been rapid and may be due to attempts to force educational research into traditional designs which are unsuitable for the questions being asked and the materials being investigated.

Ramsey and Howe (97), in reviewing research related to the teaching of science in the secondary schools, point out some of the weaknesses in these research efforts. They tentatively conclude that confusion in terminology used to describe certain instructional practises, true differences in courses and external arrangements being hidden due to the overwhelming effect of teacher characteristics; and the lack of sensitive evaluation instruments were some of the common deficiencies. On a more positive side they suggest that there is mounting evidence that instructional procedures can be designed to produce specific outcomes such as critical thinking, desirable attitude change, understanding of the scientific enterprise, as well as the more traditional objective of content acquisition.

Research dealing with the application of curriculum materials is essentially applied research. It is practically impossible to use experimental controls which would be considered adequate in the traditional sense of control group versus experimental group. In a given city, contamination between the two groups is almost inevitable. This suggests that an experimental design would have to recognize these limitations and concentrate upon carrying out an intensive investigation of conditions as they actually exist.

In a given school situation there is an interdependence of factors which can seldom be viewed in isolation. Rather, what seems to be involved is a matrix of elements. These elements may have little meaning when viewed in isolation but by their interrelationship produce a situation peculiar to the particular combination, for example, the situation of student A with teacher B for lesson C is seldom duplicated by any other student, teacher, and lesson combination. It may be possible to attach meaning to these combinations and to measure their effects by the use of observational techniques and evaluative instruments.

II. THE BACKGROUND OF THE STUDY

During the past eight years, the Edmonton Junior High School Process-Approach Science Project, under the direction of M.A. Nay has been developing a three year sequence of science courses. As reported by Nay (80) this project is an amalgamation of several features of *Science - A Process Approach*, the staff's perception of the nature of science, and the realities of science teaching.

From its inception in 1965, the project has been a cooperative research and development effort in which school administrators, teachers and university personnel were directly involved in the attempt to produce a viable curriculum which would give students an understanding of and skill in the process of science concomitant with an acquisition of science knowledge. The framework for building this curriculum is supported by the following beliefs:

1. Content or science knowledge is of primary importance and dictates the problem to be investigated and the strategies of inquiry to be utilized.
2. The structure of a given science should be emphasized and careful attention given to the organization of this science knowledge.
3. It is possible to classify the activities of

scientists into simpler activities or processes and there are general processes used by all scientists at some stage of their research.

4. The processes of science, including the intellectual activities associated with it, can be learned by students. This learning can be facilitated by means of an instructional sequence in which the students carry out activities similar to the activities of the practising scientist. Students are not miniature scientists so that the inquiry activity in the junior high science course must be determined not only by the nature of the problem being solved but also by considerations relative to how students learn.

Mokosch (76) reported that the project was initially concerned with how to translate the working methods of the scientists into procedures adaptable to investigations carried out by students in classroom situations. In other words, to take the significant processes used by scientists and make them operationally suitable to guide the investigations carried out by students.

A major step toward the solution of this immediate problem was the production of _An_ _Inventory_ _of_ _Processes_ _in_

Scientific Inquiry*, which will henceforth be referred to simply as the "Inventory".

In producing the Inventory, the members of the Edmonton Junior High Science Project used a large variety of sources such as biographies, original papers, books by science historians, observations of scientists at work, etc. and prepared a comprehensive list of activities which were relatively common in the research efforts of scientists in the various scientific disciplines. Each of the activities or processes of science in the prepared list was described and its purpose delineated. The activities were then arranged in a formal structure to aid in illustrating the interrelationship among the processes and to provide some guide to the learning process.

The versatility of the Inventory enabled it to be used to encompass the curriculum within a framework which directed the teacher to plan investigations which reflected the processes of science. The fact that most of the processes were stated in behavioral terms made evaluation and planning procedures complementary.

Hopefully then, a curriculum could be constructed which would structure the teaching of scientific topics and direct teaching activities so that the content of science would be interpreted within the meaning and context of the process which led to the acquisition of that content. This approach would conform, to some extent, to a statement made by Tanner:

---

*See Appendix A for a copy of the Inventory.

> If the structure of the discipline encompasses the modes of inquiry for that discipline, theory of knowledge and content cannot properly be separated from the teaching process. (117).

Mokosch (76) carried out an investigation of the project by comparing four treatment groups of junior high school students in an attempt to determine whether process in science could be taught by means of a curriculum operating within the framework of the Inventory and whether growth along the process dimension could be effectively measured. Although the results of this investigation did not show that the new method was superior to the traditional method, it became clear that methods of evaluation and teaching along the process dimension required further investigation.

The work of the first two years which included in-service training for participating teachers, the production of the Inventory, the application of the Inventory and the subsequent evaluation of the resulting program may be considered as phase I of the project. This writer became associated with the project toward the end of the first stage and became interested in its further development and the related evaluative aspects. At this time Mokosch's study was practically completed and, although disappointing in some respects, gave encouragement by showing that the new approach was at least as successful as the traditional one in most areas. It was reasoned, therefore, that if the methodological and evaluative procedures could be improved and refined, a curriculum could be produced which would reflect both the product and process of science as they are related within the disciplines.

Accordingly, the information obtained from field observations and from Mokosch's study were used in the continuation of the Matter and Energy program in a number of classes and in the adaptation of the Inventory to a seventh grade Life Science program. The improvement of the methodological and evaluative procedures became the concern of two separate studies. One study by David Powley (94) was concerned mainly with the methodological portion by attempting to develop appropriate learning experiences within the framework of the Inventory. The present study is concerned with evaluating the operational aspects of the curriculum and the level of achievement of the student participants. There was some overlap in these studies so that areas of joint concern were approached by means of a co-operative endeavor.

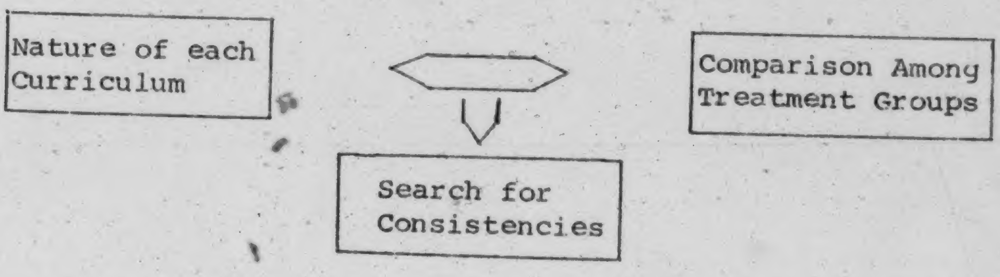
III. THE PROBLEM

This study has two main aims: (1) to develop evaluative procedures for a process approach curriculum, and (2) to use these procedures to gain information relative to the problem of how to teach science so that students will gain skill in, and understanding of, the process dimension of science.

The study is associated with the developmental aspects of a particular science program and its evaluation. Few real manipulative procedures are possible, or even very meaningful, in dealing with this particular kind of classroom research. This is essentially applied research in which data are gathered in a controlled manner in an attempt to discover significant relationships. In this sense, this study is experimental within the definition given by Cattell:

> An experiment is a recording of observations quantitative and qualitative, made by defined and recorded observations and in defined conditions, followed by the examination of data by appropriate statistical and mathematical rules for the existence of significant relations. (21)

Although a curriculum may be specified in advance, the operational aspects of the curriculum would tend to exert modifying influences which could profoundly alter the true nature of the curriculum under which the students actually operated. The problem to which this study is addressed deals with ascertaining the nature of the true curriculum and attempting to relate the true curriculum to achievement levels in the process dimension. The relationships between the main concerns of this study may be illustrated by the following model:



The curriculum refers to the educational program specified in terms of what the teacher actually had done with the curriculum materials and learning aids at his disposal. A treatment group refers to the students who were subjected to the same curriculum. Comparisons among treatment groups were made on the basis of achievement levels on a selection of tests. The results of these comparisons were subsequently related to the curriculum of each group.

This study was carried out using six treatment groups of grade seven students defined initially according to the teacher from whom they received instruction. Each teacher subscribed to the same overall behavioral objectives, as specified by the Inventory, so that the curriculum would differ mainly in factors related to teacher interpretation and implementation of the curriculum materials. The nature of each curriculum was determined from reports submitted by each teacher supplemented by formal observation procedures using an instrument designed for this purpose. Comparisons among treatment groups were made on the basis of tests prepared, adapted, or selected to measure aspects of the process dimension.*

The comparisons among treatment groups were made in order to determine whether differences appeared to be related to the form of instruction received and/or the degree to which the form of instruction was consistent with that implied by the Inventory. It was anticipated that, when the strengths and weaknesses of each treatment group was compared to the instructional procedure used, some judgment could be exercised about the relative effectiveness of techniques and methods used in teaching for process.

---

*Copies of prepared or adapted tests are to be found in Appendix C.

Definitions

Some of the following terms have already been defined when they were first introduced but they are included here with the other definitions for the sake of clarity.

Affective Attributes of Scientists refers to personal characteristics demonstrated in common by scientists as they go about their work and included in the categories indicated by the common terms "appreciations", "interests", "attitudes", "values",and beliefs", and "adjustments", (81).

Inventory refers to *An Inventory of Processes in Scientific Inquiry* developed by the Edmonton Junior High School Science Project under the direction of M.A. Nay.

Curriculum refers to the educational program specified in terms of what the teacher actually had done with the curriculum materials and learning aids at his disposal.

Experimental Curriculum refers to the curriculum dealing with biological science at the grade seven level which operated within the framework of the *Inventory*.

Curriculum Evaluation refers to all systematic efforts, including the use and interpretations of all instruments and the allied analyses, used to assess strengths, weaknesses, and usefulness of the experimental curriculum.

Inquiry refers to an active form of learning in which the individual asks, investigates, and questions as he searches for meaning in nature. Inquiry includes the actions of searching, collecting and processing of data, discovering, and verifying.

Process refers to the "doing" aspects of science and includes those procedures that scientists carry out in their search for meaning in nature. This includes those procedures as outlined in An Inventory of Processes of Scientific Inquiry and may include other scientific procedures not specifically included in the Inventory.

Co-operative Science Test Score (COOP test score) refers to the number of items a student had correct on the Co-operative Science Test - General Science - Form B.

HIFAMS score refers to the score obtained on the How I Feel About My School attitude test.

Process of Science Test score (PST score) refers to the number of items a student had correct on a test designed to measure a student's knowledge and understanding of the processes of scientific inquiry used in two filmed episodes.

Science Reasoning Test score (SRT score) refers to the number of items a student had correct on a test designed to measure a student's ability to recognize the processes of scientific inquiry and to utilize these processes in a verbally simulated situation.

TOUS score refers to the number of items a student had correct on the *Test On Understanding Science - Form eW*.

Inquiry Efficiency Test score (IET score) refers to the score obtained on an adaptation of the *TAB Science Test*.

Hypotheses

The testing of hypotheses was considered to be the preliminary phase of this particular study and not an end in itself. The main purpose was to relate treatment procedures to student abilities and characteristics as estimated from a number of instruments including those whose scores are defined above.

Comparisons among groups were made in an attempt to determine whether differences existed among treatment groups in the following areas: (1) knowledge about science and scientists (TOUS), (2) ability to engage in the processes of science as listed in the Inventory (PST and SRT), and (3) ability to select questions whose answers are most likely to lead to a solution of a scientific problem (IEF). Comparisons among groups on the basis of attitude were made mainly to provide information which could temper later judgmental decisions relating treatment procedures to achievement.

Hypothesis 1.0 There will be no difference among the means of the treatment groups in the scores on the HIFAMS attitude test.

1.1 There will be no difference among the means of the treatment groups in the scores on the Process of Science Test.

1.2 There will be no difference among the means of the treatment groups in the scores on the Inquiry Efficiency Test.

1.3 There will be no difference among the means of the treatment groups in the scores on the Science Reasoning Test.

1.4 There will be no difference among the means of the treatment groups in the pretest scores on the TOUS.

Hypothesis 1.5 There will be no difference among the means of the treatment groups in the posttest scores on the TOUS.

In accordance with the assumptions behind the construction of the various tests, it was considered essential to allow for initial differences between groups by the use of covariate control. This led to the testing of a number of further hypotheses:

Hypothesis 2.0 There will be no difference among the adjusted means of the treatment groups in the scores on the Process of Science Test using the I.Q. scores and COOP scores as covariates.

2.1 There will be no difference among the adjusted means of the treatment groups in the scores on the Inquiry Efficiency Test using the I.Q. scores and COOP scores as covariates.

2.2 There will be no difference among the adjusted means of the treatment groups in the scores on the Science Reasoning Test using the I.Q. scores and the COOP scores as covariates.

2.3 There will be no difference among the adjusted means of the treatment groups in the posttest scores on the TOUS using the I.Q. scores, the COOP scores, and the pretest scores on the TOUS as covariates.

By means of an instrument, specifically prepared and adapted to make controlled classroom observations, two judges obtained information which was used to test one further hypothesis:

Hypothesis 3.0 There will be no difference between the teacher's perception of the science process he is emphasizing and the science process he is actually emphasizing.

Delimitations

1. The sample used in this study was composed of grade seven students in the Edmonton Public School system.
2. The achievement tests used in this study were directed mainly at the process aspects of science. No implications about the relative effectiveness of the various treatments for content areas can be implied.
3. This study represents the effects of the various treatments over a period of approximately four months. Magnification or diminishment of differences could occur if the time period was extended or diminished.

Limitations

The main concern of this study dealt with conditions as they actually existed so it was necessary and desirable to use intact classes, which obviously makes any assumptions of randomization untenable. The samples were considered to approximate random samples of the population of all students (past, present, and future) of each individual teacher. Associated with applied research of this type, there is the necessity of obtaining the whole-hearted support of the school officials and participating teachers. It was therefore essential to keep testing procedures within the bounds of the normal school routine.

The considerations as outlined in the preceding paragraph point out two major limitations. First, the groups cannot be considered to be random samples of the present junior high school population and second, the test instruments which were used would form a limited sample of testing instruments.

A further limitation was the assumption that the _Inventory_ was a valid description of procedures carried out by scientists as they go about their work. This assumption was considered to be quite tenable but the interpretation of the _Inventory_ by the individual teacher was considered to be another variable which required examination. This meant that the treatment variable for each group would have to be categorized and labelled in accordance with a sampling obtained from reports and observational techniques.

III. SIGNIFICANCE OF THE STUDY

Recent attempts to produce curriculum materials seldom considered including provision for change within the curriculum itself. Some of these materials have emphasized the processes of science but after an initial testing period, almost invariably these materials crystallized into bound volumes which are difficult to revise or adapt to changing conditions.

If the use of the Inventory can be made to function so that individual initiative and creativity can operate within its framework, then it may be possible to construct a truly viable curriculum. This would mean that the curriculum would have form and structure, as determined by its framework, to direct the study of the subject matter. The "doubt component", inherent in any science course taught within the framework of the Inventory, would ensure that the subject matter or content be open to challenge. In essence, revisions and adaptations would form an expected and essential part of the curriculum.

Although there is general agreement that some science teachers are more successful than others, little is known, in terms of specifics of what teaching behaviors produce certain results in students. Further information in this area is essential in order to produce a sort of "science of science teaching" as opposed to the recipe type of methodology which is often in vogue.

Educators have criticized testing procedures which test mainly the lower order categories (as classified according to Bloom's taxonomy) but the problem of classifying test questions is practically insolvable if: "It is necessary in all cases to know or assume the nature of the examinees' prior educational experience." (13).

This writer is inclined to the view that the above quotation does not go far enough. Not only must the extent of the examinees' educational experience be known, but also their knowledge, skills and abilities as gained from all of their experiences. Because of the interdependence of the classification of test questions and the nature of the examinee's prior educational experience; no single "correct" classification, which includes the higher order categories of mental activity, would be likely to exist. A question which could be classified under the analysis of relationships for one student might be a pure knowledge question for another, and so on. It follows, therefore, that an individual teacher would experience a great deal of difficulty in preparing test questions which would assuredly test for higher order categories of mental activity.

When the process as well as the product of science is emphasized, the difficulty experienced in the classification of test questions could become even more complex. For example, if each of several students carried out a scientific investigation, one might hypothesize on the basis of previous knowledge, another by using evidence from the investigation, while a third might arrive at a reasonable hypothesis by means of an intuitive guess. As testing procedures should correspond to the learning situation, the problem of the classification of test questions may not be crucial in process type tests in which students are expected to perform and demonstrate their skills and abilities in real or simulated scientific investigations. This emphasizes the need to investigate ways and means of designing testing procedures suitable for groups in which the problem foci make use of stimuli which closely resemble real situations in a scientific investigation.

CHAPTER II

REVIEW OF RELATED LITERATURE

The literature which was considered pertinent to this study may be divided into three main classes: (1) considerations related to the rationale of the Experimental Curriculum, (2) evaluation in Science Education with emphasis on the process dimension, and (3) techniques in classroom observations.

I. CONSIDERATIONS RELATED TO THE RATIONALE OF THE EXPERIMENTAL CURRICULUM

A. Introduction

The need for reform in science education was made quite evident by the out-of-date content of high school courses of the early 1950's and, as a consequence, a number of courses such as PSSC (Physical Science Study Committee, CHEM Study (Chemical Education Material Study), and BSCS (Biological Science Curriculum Study) were developed. These courses were designed by scholars in the particular disciplines with the cooperation of teachers and were financed by the National Science Foundation (NSF) who seemed to have established the criterion that NSF financial support would go mainly to projects which were directed by scientists (84). Whatever the politics behind the funding, the scholars were concerned with the presentation of the structure of the disciplines. Ostensibly laboratory work forms a large part of these programs and the spirit of inquiry is emphasized. Laboratory programs introduce the concept of "uncertainty" and are designed to show students that the findings of scientists are subject to change. Such changes are dependent on the discovery of new evidence and could be interpreted only with reference to the precision of the instruments which were used. Science was thus presented as a search for explanations rather than a collection of known empirical truths. Essentially the scientists seemed convinced that one of the ways for a student to understand science was to have him carry out investigations in a manner somewhat similar to that of a practising scientist. For example, students in the PSSC course were expected to use simple laboratory apparatus to discover such things as the nature of wave motion, the properties of vector quantities and Newton's Laws of Motion; and George Pimentel in referring to the laboratory activity of the CHEM Study course stated: "It permits you to engage in scientific activity and thus, to some extent, to become a scientist yourself"(92).

In retrospect, these courses reflected the concern of scientists that students acquire meaning from their studies and activities in the field of science. In this one aspect, at least, the courses represent not so much a change in the basic aims of science education as a challenge to what constitutes meaning within a discipline.

This concern for meaning is not new. John Dewey (29) had emphasized the role of participation and activity in producing meaningful learning. His contention that concrete empirical experiences have meaning for children who may not understand the manipulation of a corresponding abstraction has been demonstrated by Piaget (90). Holt (52) cautions against having children manipulate symbols whose meaning they did not understand. This, Holt suggests, tends to have a negative effect and children begin to feel that all symbols are meaningless.

If education is a search for meaning, some consideration must be given to the developmental level of the students, the problems of sequence, and the importance of structure. These implications follow or appear to follow from the work of a large number of learning theorists. Of particular interest, because of the applicability to a junior high school curriculum, are the work of Piaget in developmental psychology, that of Gagné in considerations of sequence, and that of Bruner in discovery learning and the importance of structure.

B. Developmental Theory and the Importance of Structure

Piaget, although primarily a developmental psychologist, has engaged in theoretical and experimental investigations which have had profound implications for education. The invariance of stages of intellectual development is a particular theme which may have been misinterpreted. Piaget is adamant about the existence of stages but maintains that good pedagogy can have an effect on this development. This is implied, rather than being stated explicitly, in much of Piaget's writings as it is in the following quotation:

> If there are, in causality, signs of a structure which eludes empirical explanation, it will have to be admitted that this structure is plastic, and this leads us back once more to the hypothesis of an assimilation of external objects by the organism, such that the objects modify the organism, and such that the organism in its turn adapt things outside to its own peculiar structure (90).

Evidently, since pedagogy concerns itself with the provision of an environment considered to be most suitable for learning and intellectual development, the degree to which suitable "external objects" are made available must surely influence the development of the "organism". Inhelder and Piaget (55) are slightly more explicit as they illustrate by detailed considerations of the actions carried out by the subjects and the classification of these actions by inferring whether the action was modified, that although discontinuous phases may be inferred, the processes which engender these phases are continuous. Thus each successive phase is considered to be most probable because of the results of the preceding phase. Here pedagogical techniques may be brought to bear directly upon the production of results considered to be essential to advancement to a desirable successive phase.

Piaget's classification of the stages of intellectual development may be summarized as follows:

I. Sensori-motor (ages 0-2 years)
II. Pre-operational (ages 2-4 years)
III. Intuitive operational (ages 4-7 years)
IV. Concrete operational (ages 7-11 years)
V. Formal operational (ages 11-15 years)

Only the latter two stages above are of much concern to teachers at the junior or senior high school level.

Although Piaget has indicated that most Swiss children should become formal operational and be capable of abstract reasoning at between eleven to fifteen years of age, this is not necessarily true of North American children. Friot (37) found that 82 per cent of eighth and ninth grade children were still concrete operational and McKinnon and Renner (70) report that fifty per cent of a sample of students entering college were operating at the concrete operational level. It cannot be assumed therefore that Canadian junior high school students have progressed much beyond the concrete operational level.

During the concrete operational stage the child develops a form of reasoning ability. He is able to classify objects, arrange them in serial order and acquires concepts such as the principle of reversibility of operations. The reasoning ability at this stage is applied to concrete physical situations but cannot be applied logically on a purely symbolic level. This does not imply that children in the concrete operational stage lack the abilities to hypothesize or to make reasonable inferences but rather that children in this stage of development are unlikely to demonstrate these abilities in the symbolic abstractions of language or mathematics.

In a process approach curriculum, the principles involved in the more abstract reasoning of actual scientific investigations could be consciously programmed into student investigations. The actual manipulation of concrete objects in which hypotheses are generated, observations carried out, and inferences made should be central to this curriculum. Mental "structures" cannot be considered to be rigid and the assimilation of input must produce a modifying effect so that the individual can adapt closely fitting input to his own particular mental structure. The active participatory nature of a process approach curriculum makes it more probable that the need to resolve logical inconsistencies results from the child's own actions. In this manner intrinsic motivation is provided and successful resolution of the inconsistency might well represent an increment in cognitive growth.

Bruner (16) recognizes the existence of Piagetian stages of development and proposes that instruction can be designed to match the structure of the subject matter to the child's characteristic way of viewing things. He formalizes his ideas on instruction into the following four theorems:

1. A theory of instruction should specify the experiences which most effectively implant in the individual a predisposition toward learning.
2. A theory of instruction should specify the ways in

which a body of knowledge should be structured so that it can most readily be grasped by the learner.

3. A theory of instruction should specify the most effective sequence in which to present the material to be learned.

4. A theory of instruction should specify the nature and pacing of rewards and punishments. (16)

The "act of discovery" is given an important role by Bruner (14). He hypothesizes that learning through ones own discoveries could produce the following benefits:

1. The learner would acquire information selectively so that it would be more viable in problem solving.

2. The learner would achieve gratification because the act of discovery would satisfy an intrinsic need to deal with the environment.

3. The learner would generalize from the effort of discovery into a state of inquiry which would be applicable to a variety of tasks.

4. The learner, through discovering information for himself, would retain the information so that it would be more readily accessible in memory. (14)

He considers any domain of knowledge to be of such nature that it could be put into a form which would be understood by the individual learner. This form and representation would be adaptable to the individual's learning ability. It is assumed that the representation would vary from the concrete to the abstract, namely, enactive, ikonic and symbolic. Implicit in Bruner's theory are considerations from developmental psychology such as that illustrated by Piaget. The mode of problem solving would be adapted to the individual so that abstruse concepts could be presented using a representation suitable for the particular stage of development. These procedures may be questionable in some instances, but the belief, that appropriate instruction can effectively increase the number of people capable of attaining mastery of a body of knowledge, has been substantiated by Carroll (20) and Bloom (12).

Bruner's classification of the form and representation of knowledge into enactive, ikonic and symbolic categories parallels an approach which could form an integral part of a process oriented curriculum. The activity involved in the laboratory and in field trips could form an essential part of the instructional sequence, particularly for students at the concrete operational level. Actual experimentation could be supplemented by "vicarious experimentation" involving such things as the results obtained by classmates, the use of suitable case histories of science, and the use of questions of the, "What would happen if . . . ?" variety applied to procedural changes in familiar laboratory investigations. Finally, provisions must be made within a process approach curriculum to allow for degrees of sophistication in dealing with the processes. Although many students could be dealing adequately (for them) with the processes at a concrete operational level, those students who were capable of dealing with abstractions must be allowed scope and opportunity for their talents.

Bruner's conception of discovery learning is quite compatible with some degree of formalism in the instructional sequence. He interprets discovery under the following assumption:

> . . . discovery . . . is in its essence a matter of rearranging and transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights.(14)

Conceivably, the formalism suggested by the four theorems of instruction would guide discovery learning with considerations related to the structure of the subject matter, the sequence of presentation, and the selection of student investigations in which the chance of success (reward) and failure (punishment) would provide incentive to further learning.

C. Scientific Inquiry and the Process Approach in Science Teaching

Usage in the literature suggests no clear cut distinction between discovery learning and inquiry. For the sake of clarity, the term inquiry will be considered here to be similar to discovery learning in that both search for truth or knowledge but only inquiry must always be concerned with methods and techniques and the associated bounds of the ensuing knowledge. In this sense Bruner's (14) "Insightful Discovery" or Stotler's (114) "Discovery Approach" could be inquiry for some students but not necessarily for all.

In practice, discovery learning had something of a bandwagon effect with methods varying from relatively unguided situations to situations in which inquiry had direction and focus. Ausubel (8) seriously questions the use of discovery techniques and condemns them as being wasteful in some instances. Newton (83) questions whether inquiry teaching is necessarily sound pedagogy. Although he does not suggest that inquiry teaching should be abandoned, his tour of thirty five colleges in over twenty states left him with the impression that:

> If science teaching in the next decade is to be honest to the discipline, it must resist the spirit of evangelism for inquiry teaching and return to a more balanced view of the nature of science and the nature of the learner and their implications for science education. (83)

Schwab (104) would agree at least partly with Newton's observations but he does see merit in the provision for practice in inquiry in the laboratory or classroom. Gagne (40) sees some limitations to the value of such practice and suggests that practice in inquiry prior to the student's acquiring of adequate prerequisites is not likely to be beneficial.

Suchman (115) indicates that inquiry may be divided into the actions of searching, data processing, discovery, and verification. In the Inquiry Training Program, he and his colleagues attempted to relate the foregoing actions of inquiry to a technique which would make the individual more autonomous in inquiry. Suchman attempted to assess student growth in the ability to enquire by using techniques related to the verbal interaction between the teacher and the pupil. Although some of the processes of scientific inquiry correspond to the "actions of inquiry", the Inquiry Training Program tended to concentrate on verbal behavior associated with selected investigations. In spite of the limitations inherent in the program, further evidence was added to the following beliefs:

1. Children do not automatically discover knowledge which leads to understanding, i.e. some "stage setting" by teachers could be beneficial and even necessary at early stages.

2. It may be possible to train students to become more proficient in inquiry.
3. Although inquiry behavior may be complex, efficiency in facets of inquiry may be approximately measured.

Schwab (104) advocates an approach to science teaching in which its materials would exhibit science as enquiry and would lead students to enquire into these materials. He suggests the use of materials such as: (1) original research papers translated into idiomatic expression where necessary, (2) "Narratives of enquiry" in which the data and interpretations of the scientists are described so that the conclusions of the enquiry are illustrated as formulations of the evidence possibly leading to further enquiry, and (3) "Invitations to enquiry" which pose problems relating to the subject matter under study in order to supplement the subject matter coverage by questioning unexplained terms and hints of dogmatic assertion.

Although Schwab advocates an enquiring classroom, he denigrates two tendencies associated with some of the process oriented curricula. One tendency is to convert science classrooms into miniature research laboratories where techniques are learned and data collected, yet the problem under inquiry is treated as something given and the problem of interpreting the data is avoided. The second tendency is to divorce process and content and "treat them both as orthodoxies by a rhetoric of conclusions" (104). Schwab's view of inquiry may be summarized by two of his own statements:

> 1. Treatment of science as inquiry consists of a treatment of scientific knowledge in terms of its origins in the united activities of the human mind and hand which produced it; it is means of clarifying and illuminating scientific knowledge.
>
> 2. Each (inquiry) arises in relation to a specific subject matter and the essence of each lies in its own substantive conceptions, its own data, and its own questions asked and answered. (104)

Here Schwab is in agreement with Ausubel who condemns the two practices of assuming that students can learn science effectively by playing the role of "junior scientist" and suggesting that there is a method of scientific inquiry applicable to all scientific disciplines.

Parker and Rubin (88) See process as content and argue that process should be the central link in curriculum design. They consider that knowledge, when linked to process, would be dynamic and that the elements of process would be vital to the learning situation. The requirements for a process based curriculum, as seen by them would:

> . . . deal primarily with the identification of worthwhile processes to which students should be exposed, the design of instructional strategies that make effective use of the processes, and the realignment of subject matter so that it complements the instructional strategies. (88)

Although they insist that each of the requirements reflect the character of the discipline or disciplines under which it operates, their point of view differs from Schwab's, not only in terminology but in fundamental approach. Schwab sees process and content as interdependent. Each inquiry (and hence the processes related to this inquiry) would arise in relation to a specific subject matter. Parker and Rubin would select processes into which subject matter, retooled if necessary, would be fitted. Neither Schwab or Parker and Rubin (88) make much reference to implications from learning theory although Schwab does suggest that learning theory was not sufficient because of the nature of modern disciplines. Gagne (42), however, related his learning theory directly to an analysis of the processes of science as he assisted the American Association for the Advancement of Science in producing what appears to have become a relatively successful elementary science curriculum.

Gagné (4) has effectively used task analysis as the basis for providing efficient training situations. He emphasizes the importance of structure, but considers that the best known principles are inadequate in dealing with training situations. It is generally conceded that learning involves greater complexities of cognition than does training. Glaser (44) differentiates between training and education. Training, he maintains, tends toward a specific competence and specific objectives as opposed to the broader objectives of education. In one sense, training and education are opposites in that training seeks uniformity while education attempts to maximize individual differences by discovering and releasing individual potential.

Gagné (40) devotes an entire volume to set out his views on the conditions of learning. Eight kinds of learning, forming an hierarchy from simple to complex, are identified as follows: signal learning, stimulus-response learning, chaining, verbal-associative learning, multiple discrimination, concept learning and problem solving. He admits that there may be other categories but his classification involves the proposition that transformation procedures may be designed in order that a student may move from one level to a higher level.

Gagné suggests the possibility of planning a sequence of instruction within content areas. This would be done by analyzing a given objective of learning in order to determine what its prerequisites must be and then mapping out a sequence through which the student would attain mastery of the prerequisite capabilities. Alternate routes would be allowed but the mapping procedure would ensure that no prerequisite capabilities would be omitted.(40) The foregoing procedure is readily recognizable in the AAAS program which has the following five basic assumptions inherent in its design:

1. The behaviors of scientists at work, however complex, are analyzable into simpler activities.
2. The processes of science are highly generalizable across scientific disciplines.
3. The processes of science may be learned and it is reasonable to begin with the simplest ones and build the more complex ones out of them.
4. A reasonable sequence of instruction in process skills may be designed for children beginning with the simplest ones and continuing in hierarchial fashion to the most complex.
5. At the end of the instruction the student will have acquired process skills which are broadly transferrable across scientific disciplines.

The AAAS program called Science - A Process Approach attempts to build process or inquiry skills in an hierarchial or sequential manner but content seems to be incidental to process. Whether structurally organized knowledge will be acquired efficiently with this approach may be open to some question but Gagné recognizes the importance of this type of learning as it relates to further learning as he states:

> To be an effective problem solver, the individual must somehow have acquired masses of structurally organized knowledge. Such knowledge is made up of content principles, not heuristic ones. (40)

Gagné (40) considers that the ability to engage in scientific enquiry is one of the most essential goals of science instruction but considers that certain kinds of performance capabilities or competencies are fundamental to learning about science. It would appear that the somewhat incidental nature of content in the AAAS program stems largely from the belief that competency in process skills such as observing, classifying, quantifying etc. are prerequisites for acquiring an understanding of science.

The AAAS program, as influenced by Gagné's learning theory, may represent an advance to a sort of instructional technology. Objectives are defined in operational terms, processes are broken down into basic skills and arranged in hierarchial order with their inter-relationships illustrated, and testing procedures form an integral part of the program. The organization of the program may thus be the source of its main strengths and could well contribute to some possible weaknesses. On the positive side, the development of process skills and abilities are not left to chance and students are given an opportunity to develop and practice these skills by participating in laboratory experiences. On the other hand, the use of processes usually from simple to complex rather than in terms of need for a particular investigation would do little to acquaint students with the nature of science. Associated with this slight tendency of the AAAS program to separate process from meaningful content there is some tendency to depend on the satisfaction achieved from acquiring skills as practically the sole basis for intrinsic motivation.

D. The Edmonton Junior High School Process-Approach Science Project

The Edmonton Junior High School Process-Approach Science Project was initiated in 1965 to develop a three year sequence of science courses. As reported by Nay (80) this project is based on the following four major beliefs:

1. Science knowledge or content is of primary importance in the science curriculum and dictates the problems to be investigated and concomitant strategies of inquiry to be utilized. This is the position taken by Schwab (103) and content is never used simply as a vehicle for the learning of the processes of scientific inquiry.
2. Bruner's contention that ". . . subjects have a fundamental structure that is basic to effective learning and must therefore be central to teaching.", is a sound idea. Attention must be directed toward

the emphasis of the fundamental concepts of a given science and the organization of this science knowledge.

3. The activity of scientists can be broken down into simpler activities or processes. There are general processes common to the work of all scientists but each of these processes is applied only as dictated by the problem being researched in the discipline.
4. The intellectual activities of scientists can be learned by students and an instructional sequence must be designed for this purpose. This belief is in agreement with that of Gagne (42) but the inquiry activity in the junior high school courses is dictated by the nature of the scientific problem being studied and tempered by factors related to the developmental level of students and the educative process.

Mokosch (76) adopted the postulate that it was possible to integrate the processes of science into a science course by superimposing the Inventory upon manageable units of the science content. Mokosch hypothesized that students who studied science content into which the processes of science, on the basis of the Inventory, were consciously integrated, would achieve greater growth in the process dimension than students who studied science in the traditional manner.

Mokosch studied this problem using a sample (N=668) divided into four treatment groups as follows:

Treatment A (N=172) Students in this group were taught by teachers who were familiar with the Inventory and its purpose and used prepared materials in which content and the associated processes were integrated.

Treatment B (N=213) Similar to Treatment A except that teachers in this group received no special instruction relative to the use and purpose of the Inventory.

Treatment C (N=213) This group may be considered to be the control group for the first two experimental Groups. This group was taught the same content as Groups A and B but teachers taught the material in their own way.

Treatment D (N=142) This group was apparently included mainly to check for effects of testing and maturation. Content covered by this group was only incidentally, if at all, related to administered tests.

The evaluation was carried out using a series of film loops followed by multiple choice items designed to measure student understanding of the processes of science. The film loops were graded into three categories containing three film loops each as follows:

I. Filmed episode depicts processes of scientific enquiry used in an investigation actually conducted in class.

II. Filmed episode depicts processes of scientific enquiry in a situation quite similar to that performed in class but not identical to it.

III. Filmed episode depicts processes of scientific enquiry in a content area not directly related to area under study.

Each set of film loops and the associated multiple choice items formed a process measure test whose score was simply the number of items a student answered correctly.

Mokosch used the pretest scores on the total process measures and the COOP Science test scores as covariates to obtain the following somewhat anomalous result:

> The group receiving no particular process emphasis in their instruction attained a higher degree of understanding of processes of science, as measured by the total process measure, than the groups which specifically concentrated on the processes of scientific inquiry. (75)

This result would suggest that either the Process Measures used were invalid or the traditional group received instruction which was superior to the instruction received by the experimental group. The latter alternative is not to be rejected because, as far as is known to this researcher, no best way of teaching the processes of scientific inquiry has ever been devised. The teachers in the experimental group were relatively inexperienced in a new approach and possibly unsure of themselves. This may have had some negative effect on their students. The traditional group had no such disadvantage as the teachers had complete freedom to devise and adapt instructional methods which they were convinced were suitable and to use procedures which they had previously found to be effective. As with many studies in which an experimental group is compared to a control group in regular classroom settings, the teacher variable may have masked, or hidden completely, the true significance of incorporating the processes of science into a science curriculum.

The question of why one form of instruction is apparently superior to another may have no immediate solution. Aikenhead (2) reviewed research findings in which the criterion instruments used measured learning about science and scientists as distinct from learning about specific subject matter, principles concepts and skills. He reported that most research reports found no significant difference between experimental and control groups that could be attributed to the science course, supplementary materials, or teaching strategies. Stephens (111) who documented his position with reference to many educational variables such as size of class, qualities of teachers as rated by supervisors, student-centred vs. teacher centred approaches, ability grouping, etc., and studied research reports and summaries of research concluded that practically nothing makes any difference in the effectiveness of instruction. Stephens postulates a theory of spontaneous schooling in which there are two kinds of forces one related to the background of the student and the other to the role of the teacher which accounts for most of the learning which takes place. This researcher is inclined to the view that variations applied to a teaching-learning situation by any outside agency must not only modify but be modified by the initial situation. It would seem reasonable, therefore, to attempt to determine the kind of instruction actually taking place rather than assuming that an applied variation would predominate and remain the determining factor in the instructional program. It is possible that Mokosch's classification of teachers into experimental and traditional groups acted as an external variation which did not, in fact, describe what actually took place within classrooms.

The members of the Edmonton Junior High School Process-Approach Science Project were not deterred by the somewhat unexpected results of Mokosch's research. They were convinced that their approach was logically sound but required some improvements in devising more successful methods of incorporating process and content within the classroom and further evidence to suggest ways and means of evaluation within the classroom.

The use of the Inventory was extended to sixteen biological science classes at the grade seven level. As had been done in the "Matter and Energy" program (76), each teaching unit was organized around an "Investigation" into which the processes of science, as dictated by the substantive content, were programmed.

A Life Science Committee prepared materials made up of a teacher's guide and a students' section. The teacher's guide indicated the concepts to be developed, provided suggestions for the utilization of the processes of science for each investigation, and listed reference materials and teaching aids considered to be suitable. The students' section complemented the teacher's section and included outlines, instructions, and questions directly related to each investigation.*

When this researcher began his study of the experimental curriculum described above, only Chapters I and II of the materials produced by the Life Science Committee had been completely prepared. These materials were not to be considered as inflexible dogma but simply the consensus of a committee of select teachers as to what could constitute a successful approach to the integration of process and content within a biological science curriculum. This outlook is expressed by David Powley as follows:

> The committee has therefore taken some of the experience gained in the _Matter and Energy-A Process Approach_ and applied them to a Life Science Program. Though there are probably other ways of doing this task, this one put forth by the committee is for your use and evaluation. It should be stressed that the teacher must also become part of this whole process and behavioral change.
>
> Teachers using this will find students' pages and a teaching guide. The teaching guide, it is hoped, will convey some of the methodology that may be attempted. (93)

The idea of bound volumes constituting an effective science curriculum was contrary to the envisaged use of the _Inventory_. The materials* prepared for use in the project were considered to be available to be added to, or subtracted from, according to the judgment of the user, providing each adjustment would not give students a false view of the scientific enterprise and the work of the practising scientist.

---

*See Appendix D for sample copies of prepared materials.

Although it is clear that variation and flexibility formed an expected part of the experimental curriculum, teaching within the framework of the Inventory meant the acceptance of two major principles:

1. The teaching of a discipline must be reflective of what goes on in that discipline and also take into account what children do when they learn it in school.
2. The students must first know the processes of scientific enquiry, and then be given a variety of opportunities to use them consciously in learning science. (79)

The principles as enunciated above require some further clarification. Although the teaching of a discipline must be reflective of that discipline, it is understood that children are not miniature scientists. It would be extremely unlikely that many junior high school students would have sufficient patience and drive, with the associated capability for sustained effort, to carry out independent investigations over an extended period of time.

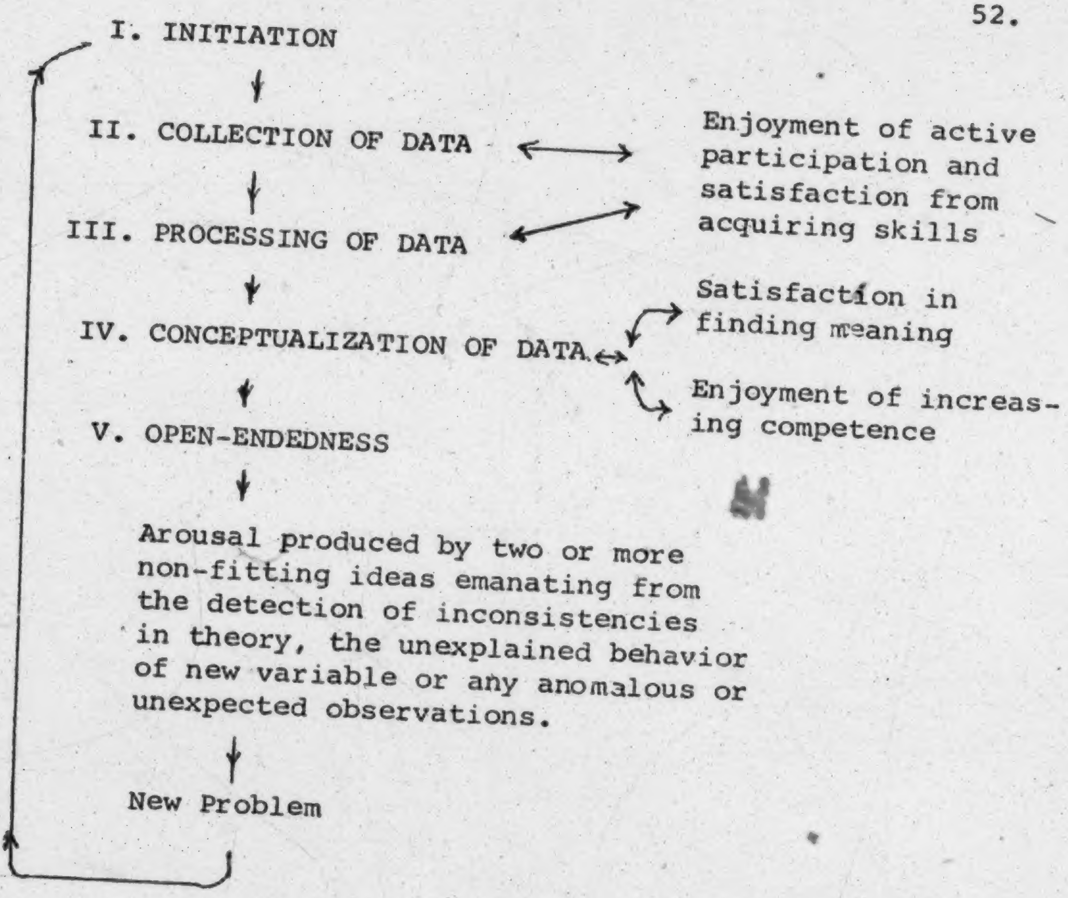
This would imply that, for most students, the "goal" should appear to be within reach and most tasks would give some promise of culmination within the temporal interest span of each student. Initially, structuring would be essential for the majority of students with the degree of structuring decreasing as the capability for more independent action increased.

The Inventory (see Appendix A) serves as a useful formalism to provide structuring within which the "investigation" could take place. The major processes, as listed from one to seventeen, follow the order in which the processes of science could be utilized in some investigations but the number of processes used and the order of their utilization would be dependent upon the subject matter of each investigation. It is the task of the teacher to decide the amount of instruction required for the class and for each individual and to determine the most opportune time for each student to move on to more independent activity. This would be done by providing for degrees of structuring and guidance which would determine what Schwab (103) classifies as "levels of inquiry." The Inventory is so ordered that degrees of independence in "levels of inquiry" are introduced by omitting processes in the instructions to the children (from top to bottom in the Inventory).

Although the students are expected to get to know the processes of science and then be given the opportunities to consciously use them in learning science, the gaining of knowledge of the processes cannot be separated from the use of the processes. Knowledge of the processes of science, in the sense used here, means knowledge of the methods of inquiry implied by the _Inventory_, with or without knowledge of the specific literal referents. It is the opinion of this writer that the ability to carry out an independent investigation in science would certainly be possible without knowing the names of the processes of science, as named in the _Inventory_; and conversely, the _rote_ memorization of the names of the processes would contribute little to a student's ability to use the processes of science in actual investigations.

One of the most compelling aspects of the _Inventory_ is its tendency to follow into an almost cyclic pattern. This pattern follows from the framework of the _Inventory_ under which the curriculum operates. The five main divisions of the Inventory are listed below to illustrate.

Intrinsic motivation is built into the pattern by providing for active participation leading to satisfaction from acquiring skills and making use of what Suchman (116)


I. INITIATION
II. COLLECTION OF DATA
III. PROCESSING OF DATA
IV. CONCEPTUALIZATION OF DATA
V. OPEN-ENDEDNESS
Enjoyment of active participation and satisfaction from acquiring skills
Satisfaction in finding meaning
Enjoyment of increasing competence
New Problem

Arousal produced by two or more non-fitting ideas emanating from the detection of inconsistencies in theory, the unexplained behavior of new variable or any anomalous or unexpected observations.

considers to be a fundamental human activity and the chief motive for most enquiry, namely, the pursuit of meaning. Conceivably, there could be a progressive elevation in the sophistication of the new problems or the selected investigations to keep pace with and to encourage student development. There could, however, be some danger that an injudicious handling of the curriculum by the teacher could lessen its effectiveness in some instances. For example, if the task posed for the child is beyond his developmental or ability level, negative motivation could result. Alternatively, the use of simplistic investigations, over and above those needed to illustrate a method of enquiry could fail to effect any sustained interest. Arnold Arons made the following remark about elementary school science curricula which could very well apply to the Experimental Curriculum:

> But these materials are by no means "teacher-proof." They can be handled effectively only by teachers who hold a deep enough understanding of the subject matter to possess the security and flexibility to lead investigation rather than to dictate end results; to accept incorrect suggestions or hypotheses, recognize the misconceptions in the child's mind, and guide him into revision and correction of his ideas rather than rejecting them by assertion and insisting on a memorized conclusion. (6)

The degree of success of the experimental curriculum could thus be greatly affected by how the teacher interpreted the curriculum. The remainder of this chapter deals with methods used to evaluate process type objectives and with techniques used to sample teaching behavior for the purpose of making inferences about the categorization or analysis of this behavior.

II. EVALUATION IN SCIENCE EDUCATION WITH EMPHASIS ON THE PROCESS DIMENSION

A. Early Attempts at Measuring Performance Competencies

Process evaluation of a sort has long been associated with the laboratory aspects of science teaching

(95). The form of evaluation depended largely on the predilection of the individual instructor and included such procedures as assigning grades to some form of production which was assumed to reflect student proficiency in laboratory investigations. Common indications of this proficiency were grades assigned to such things as laboratory notebooks and reports, experiments done by students and graded by the instructor or his assistants and paper and pencil tests designed to test the objectives of laboratory work. Examples of these procedures may readily be found in many of the high schools and the under-graduate sections of the universities of Western Canada.

Kruglak (65) notes that although there appears to be general agreement on the broad and traditional objectives of laboratory teaching, the problem of measuring the outcomes of this teaching has been far from formulated, let alone solved. He points out the necessity of preparing objectives which could be put into behavioral terms, and contends that paper and pencil tests are not adequate substitutes for performance tests using real apparatus. While it is true that the skill aspects of processes cannot be adequately tested without the use of real apparatus, there are many processes as listed in the Inventory which are amenable to written tests, e.g. organizing the data, interpreting the data, formulating operational definitions etc.

The Physics Department of the University of Minnesota instituted performance tests as a regular part of laboratory work some twenty-seven years ago (124). These performance tests contained 8 "short items" and 3 "long items." Each test location was provided with a list of the apparatus and a statement of the problem to be done. To ensure uniform grading, a key for scoring was provided which gave specific instructions for scoring each individual item and general instructions applying to all items. Wall, Kruglak and Trainor (124) indicate that considerable subjective judgment enters into the scoring of performance tests and suggest that more work must be done with them before definite conclusions can be drawn about their use as evaluation instruments.

Kruglak (63) made an intensive study of laboratory achievement tests in the Physics Department of the University of Minnesota. He proposed a list of tentative objectives which he and Trainor used as a guide to prepare some 600 test items to parallel the most widely accepted instructional objectives of general physics. These items were submitted to four college physicists for rating and review in order to check for content validity. Performance tests were constructed from the items considered suitable and administered to physics students enrolled in the appropriate subject matter area so that all students enrolled in general physics at the university wrote performance tests. Kruglak compared the scores on the performance tests with such measures as high school rank, scores on standardized tests and final grades in physics and determined the product-moment correlation coefficients among variables. He also carried out an item analysis, estimated the reliability of scoring and the reliability of the performance tests. On the basis of his analysis he tentatively concluded that performance tests can be scored reliably, tend to differentiate between students in accordance with their theoretical and experimental background and measure instructional outcomes other than those measured by conventional achievement tests (63). The relatively high correlation coefficient (52) between Owens-Bennett Mechanical Comprehension Test Form CC and one of the performance tests plus the fact that all performance tests correlated significantly with this test suggest that performance tests appear to be related to understanding of the mechanical relationships involved in laboratory experience.

In an attempt to simplify the measurement of laboratory achievement Kruglak (64, 65) attempted to convert performance tests into paper and pencil tests of two forms: essay and short answer, and multiple choice. Preliminary forms of the test were administered to some 160 elementary physics students and the results analyzed. Although it was found that the paper and pencil tests were poor substitutes for performance tests, it was concluded that paper and pencil analogs of very specific skills might be successfully constructed. It was emphasized that the real situation should be accurately portrayed using such things as photographs, three-dimensional drawings, models, etc. Kruglak's performance tests and their paper and pencil analogs represent process-type tests which measure some aspects of the process dimension. Materials or representations are given, a problem is posed and students are required to demonstrate their proficiencies in processes corresponding to processes found in the following three of the five major divisions of the Inventory: Collection of Data, Processing of Data and Conceptualization of Data.

A number of tests have been developed which measure, or purport to measure, skills or abilities associated with a trade or profession (5). These tests in the main attempt to simulate real situations and the individual is judged on the basis of performance. Hubbard (53) reports a test in medicine in which a set of symptoms is given followed by a set of possible treatments. The individual makes an irrevocable choice by erasing a black mark masking the information about the results of the selected treatment. In this manner the problem unfolds and procedures may be judged on the basis of the effect on the "patient". Hubbard's test is definitely a process test and because it is administered to individuals who possess a high degree of literacy, the verbal nature of the tests would be of little disadvantage. It is conceivable that this type of test could use other form of stimuli such as sound tracks with associated pictorial representations.

More recently several tests have been developed which purport to measure achievements in science other than achievement related simply to content acquisition or laboratory proficiency.

B. Tests which measure knowledge About Science and Scientists

The Test on Understanding Science developed by

Cooley and Klopfer (23) measures knowledge about science and scientists. The form W is a four-alternative sixty item multiple choice test and elicits responses dealing with the roles of science, scientists and the scientific method. The items are categorized into three sub-scales:

I. Understanding about the Scientific Enterprise	(18 items)
II. The Scientist	(18 items)
III. Methods and Aims of Science	(24 items)

Claims for content validation were obtained by basing the items upon an analysis of scientists at work and a wide variety of literature including biographies of scientists, writings by scientists, histories of science and the philosophies of science. Professional scientists, science teachers, and professors of the history and philosophy of science were consulted regarding the content of the items to provide a further check on the content validation. After field trials and revision, the form W was standardized from data obtained from students in grades 10 to 12. The TOUS form W has a Kuder-Richardson formula 20 reliability coefficient of 0.76 and may be considered suitable for use with senior high school students (18). The TOUS form Ew is suitable for upper elementary and junior high schools grades. It contains thirty-six items and, according to communication received by this writer from Leo Klopfer, this test is still in the experimental stage. Mokosch (75) used this form of the test in his research and obtained a Kuder-Richardson Formula 20 reliability coefficient of 0.83 using pretest data and reported a Kuder-Richardson Formula 21 reliability coefficient of 0.70 obtained from a personal communication from the Elementary School Science project at the University of Illinois.

The TOUS has been used extensively as a research tool but some criticisms of it have been made. Welch (124) suggests that form W might be improved by revising some of the items and obtaining stronger evidence for content validity. Aikenhead (2) suggests that some of the items evoke a response of attitude. Hukins (54) included the TOUS in a factor analytical study of instruments designed to measure the objectives of science teaching and found that TOUS loaded strongly on a verbal factor. Certainly TOUS can be improved and being a verbal test which elicits responses relating to science and scientists, it might be expected to load on a verbal factor and be influenced by a student's partiality for the subject field. In spite of minor shortcomings, TOUS has been a useful instrument for measuring the TOUS version of "understanding science" which definitely includes knowledge about science and scientists. Although not a process test per se., it contains elements of process in that it requires the student to make judgment relative to the purpose of scientific investigations, the interpretation of data, the collection of data, the meaning of hypotheses, the formation of scientific theories, etc.

The Facts About Science Test (FAS) (112) was developed by the Educational Testing Service in 1958. It contains seventy-eight multiple choice items consisting of three alternatives each. It is scored on the basis of three scores as follows:

I. Understanding of Science as an Institution
II. Knowledge of Scientists as an Occupational Group
III. Total

The Facts About Science Test is similar in content to the TOUS sub-scales I and II. Content validity is claimed for the test but not documented (113). Cossman (26) used the FAS in a research study and found that the achievement of the control group, as indicated by the average FAS score, dropped considerably. Cooley and Bassett (22) were unsatisfied with the reliability of the FAS, as suggested by its use in their research, and concluded that an improved instrument was necessary. The FAS, as illustrated above, has not demonstrated its usefulness. It would appear that the failure to subject preliminary forms of the test to large scale field testing and subsequent revisions may have made this test unreliable and of questionable validity.

The Science Process Inventory was developed by Welch in 1966 in an attempt to produce: " . . . a valid, reliable and usable instrument to inventory the knowledge of the processes of science possessed by secondary school pupils.", (125). Form C of the Science Process Inventory consists of 150 statements to which students are asked to express agreement or disagreement. Scoring is done by counting the number of responses in agreement with a standard key.

Content validity is claimed mainly on the basis of using the judgment of experts as a guide in revising an initial list of items about " . . . the assumptions, activities, products, and ethics of science" as obtained from the literature. Further evidence for validity was obtained from its ability to distinguish among three groups composed of scientists, high school students, and high school teachers respectively. Statistical information available based on the scores of 1283 high school students gives a reliability estimate, based on analysis of variance, of .79. A revised form of the Science Process Inventory (form D) contains 135 items (126). The Science Process Inventory has been used in conjunction with other measures as part of the evaluation process in the development of Harvard Project Physics. Statistical information obtained from physics students' scores indicate a Kuder-Richardson Formula 20 reliability coefficient of .86 (127). The Science Process Inventory seems to be more comprehensive than the TOUS in dealing with the TOUS sub-scale III, The Methods and Aims of Science, but does not deal to any extent with the two other sub-scales of the TOUS.

The Wisconsin Inventory of Science Processes (WISP) was constructed by the Scientific Literacy Research Center (105) and is almost identical to the Science Process Inventory in terms of content but contains only 93 items. It requires three responses (accurate, inaccurate and not understood) as compared to the two responses of the Science Process Inventory. The scoring key (106) indicated that "inaccurate" and "not understood" are to be lumped together as simply the opposite of "accurate." This device is probably used to encourage a student to proceed to the next item rather than dwell on one which he finds difficulty in understanding, but may mislead the student into believing that there are, in fact, three separate responses. The WISP and the Science Process Inventory deal mainly with knowledge about the Methods and Aims of Science and test knowledge about the processes of science but fail to test how a student would use the processes of science in a real or simulated scientific investigation. Both tests are verbal and subject to possible misinterpretation by the student due to limitations on the part of the student, or limitations inherent in the test as a result of removing statements from context and attempting to express some processes in sentences easily understood by high school students.

C. Tests Which Measure the Ability to Use the Process of Science

The Processes of Science Test (POST) (11), initially called the BSCS Impact Test (46), was developed as part of the BSCS testing materials. The test was designed to:". . . measure understanding of scientific principles and methods with minimal reference to specific content." (67) The test consists of 40 four-option multiple choice items. Each item poses a question or situation by means of statements or paragraphs, often supplemented by such things as numerical data or graphs, in an apparent attempt to simulate aspects of a scientific investigation. Data obtained from the use of the revised test during the 1962-63 year give a pretest mean of 22.3 with a standard deviation of 6.1 and a posttest mean of 26.6 with a standard deviation of 6.5.

Mallison (67), in reviewing the test, is severely critical of it, suggesting that some of the items are little more than a measure of general intelligence and the overall test is of questionable validity as a measure of students' understanding of science and its methods, and invalid as a measure of knowledge of biology. He ends his criticism of the test with the words: "It seems to be good for everything in general, but nothing in particular." Mallison's strong criticism seemed unwarranted. While it is true that the attempts to simulate science processes could be improved by the use of problem foci which included more of the visual dimension as opposed to the somewhat excessive verbalism, the fact that the test is not greatly dependent on prior science knowledge could make it a better discriminator in the process dimension. In its present form the test has not been established as a valid process instrument but with some revisions and further evidence for validity, it could possibly become a useful instrument.

The Science Process Instrument (3) was developed by the American Association for the Advancement of Science as an instrument to evaluate the long-range outcomes of Science - A Process Approach (4) Curriculum. The instrument consists of eight sections each corresponding to a process identified by the Science - A Process Approach curriculum. Each process measure is made up of a series of tasks corresponding to levels of behavior in a process hierarchy. The successful behavior which a child exhibits on each task within each process measure is related to levels within a process hierarchy. This relationship should enable the teacher to identify the behaviors related to each process which the child has acquired and the ones which he has yet to acquire. This very direct connection between teaching and testing should be a great advantage to any curriculum. Unfortunately, the Process Instrument requires some 330 minutes to administer to each child. Even though not all children will complete every task in each measure or take the entire instrument, the procedure is so time consuming that its practicality may be questionable. Besides the obvious disadvantage of being very time consuming, there are several other disadvantages. The direct contact between examiner and examinee could introduce some bias because of some personal relationship between them; i.e., a child may react positively to one examiner but not to another. The capability of carrying out a complete scientific investigation using appropriate processes is not measured by the instrument because skill in the use of separate processes does not, of necessity, measure overall competency in the process dimension. Finally, because of the free response element and the practical impossibility of preparing a guide to explicitly specify the range of all acceptable responses, the examiner may be forced to make some subjective judgments. In spite of its limitations the instrument has minimized the effect of judgmental errors by putting the grading on a system requiring only two decisions (acceptable, not acceptable). The instrument minimizes the effect of verbal skills by utilizing oral discourse, demonstrations by the tester and actual objects. It is definitely a performance measure which measures process skills in terms of the AAAS classification.

The Questest was developed by Suchman (115), in an attempt to evaluate individual inquiry processes of children. In this test film loops depicting problem episodes are shown to children who ask questions which the examiner answers with a "yes" or "no" response. The entire session is tape recorded and analyzed on the basis of the number, frequency, and types of questions asked in an attempt to relate the child's questioning techniques to the actions of searching, data processing, discovery, and verification. A paper and pencil test consisting of three parts (Test A, Test B, and Test C) is used in conjunction with the testing procedure. The first part of the test (Test A) is administered prior to the filmed episode and the complete test administered following the questioning session. Test A is designed to measure what principles the child has discovered during the session, Test B to measure which of the necessary conditions he can identify and how accurately he can identify them and Test C is designed to measure the amount of factual knowledge of the parameters (objects, conditions, and events) of the problem episode which the child has positively identified. As an overall measure of conceptual growth, Suchman uses a test designed by him and James L. Feljar. This test uses pictorial representations with multiple choice items. Although designed specifically as a measure of inquiry behavior, it is just this particular aspect of the measures which may be questionable. In a study reported by Renne et. al. (99) only 53 per cent of a sample of 301 students asked 1 or more questions in each of three Suchman type enquiry sessions. Obviously, it would be impossible to evaluate the inquiry behavior of students who asked no questions. The remainder of the testing procedure may have some very real value in suggesting ways and means of estimating student competence in such processes as observing, collection of data, hypothesizing, prodicting, inferring etc. The use of film loops has several advantages. Film loops can be made to closely simulate actual investigations, may be presented to complete classes, and are unchanged upon repetition. Pictorial representations, although not as suitable for representing action, may be closely studied and thus serve to reduce excessive verbal exposition.

The Process Measures developed by Mokosch (76) use the film loop format to present the problem foci. This test is described earlier in the chapter. The verbal aspects of the process measures may present some limitations. Students who are familiar with the names of the processes as used in the Inventory would have an advantage over those unfamiliar with the names of the processes. It is also conceivable that a student may be capable of hypothesizing, inferring, etc. yet be unable to indicate this on a written test.

The TAB Science Test was developed by Jones (57) as an instrument to measure inquiry behaviors of students in elementary science classrooms. Jones sets up a model of inquiry behavior consisting of the five activities of searching, processing data, discovering, verifying, and assimilating-accommodating and develops a test to sample the behaviors involved in the model. He adopts the Suchman (115) technique of using a film loop format for the problem foci but restricts the "enquiry" to a set of pre-selected questions which are answerable by a "yes" or "no" response. This restriction, while limiting the range of student questioning, simplifies the analysis of student responses. The TAB Science Test consists of two problem foci presented in film loop format and contains three sections related to each film as follows:

Section I. Student selects an hypothesis from a given set.

Section II. Student selects clue questions in any given order and obtains "yes" or "no" answer by removing a laminated tab.

Section III. A repetition of the hypotheses given in Section I each followed by tab covering "yes" or "no" response.

Two multiple choice items based on a pictorial representation are used to sample the ability to transfer concepts.

The TAB Science Test is an experimental instrument whose validity rests mainly on the model set up by the author. It was subsequently used by Raun and Butts (99) in conjunction with a battery of tests and labelled as a science problem solving test. Scores on the TAB Science Test were not significantly predicted by performance on the AAAS Process Instrument. Although the reliability and validity of the TAB Science Test are questionable, it is reasonable to infer that the student who selects an hypothesis and "homes in" on the correct solution with a minimum number of steps has demonstrated efficiency in this procedure.

Mary Alice Burmester (17) developed a test which purported to measure some of the inductive aspects of scientific thinking. The test was designed specifically for a course in biological science at the college level. It was developed from a set of preliminary tests which resembled, in the areas of their concern, some of the processes of science as set out in the Inventory. A list of these concerns include the following:

1. Ability to differentiate phases of thinking. (This

includes the ability to recognize problems, hypotheses, experimental results, non-experimental observations and conclusions or inferences.)

2. Ability to delimit a problem.
3. Ability to recognize faulty experimental procedure.
4. Ability to organize data.
5. Ability to understand the relationship of facts to the solution of a problem
6. Ability to interpret data and to understand the results of experimentation
7. Ability to make conclusions
8. Ability to understand the assumptions underlying conclusions

Clearly number 1 (above) deals with the recognition of the processes and the remainder deal with abilities to deal with the processes. The test is mainly verbal, supplemented by diagrammatic illustrations and data tables for some questions, so the ability to perform in the process dimension ~~is~~ inferred rather than measured directly.

Burmester (17) claims validity for the test because its correlation with psychological examinations (reading ability, intelligence, factual information) is not excessively high, ranging from .40 to .51.

Test scores of students correlated significantly with ratings made by competent judges and students who had completed three terms of Biological science scored significantly higher than another group who had not yet taken Biological Science. Reliability as calculated using the Kuder-Richardson formula was .89. Kaplan (59) administered the Burmester test informally to Ph.D's in science and reported that there was slight disagreement on some 15% of the items but the Ph.D's scored significantly higher than the students. He also compared student performance on the Burmester test with performance on the Watson-Glaser Critical Thinking Appraisal. After one semester of biology there was significant improvement in the scores on the Burmester test but not for the Watson-Glaser Test. The evidence seems to indicate that the Burmester test is quite reliable and does measure the ability to understand and use some of the processes of scientific inquiry. It is a college test and not suitable for junior high school students.

Flynn and Munroe (36) carried out an evaluation of Nuffield Science in the secondary schools in Auckland, New Zealand. Their purpose was to " . . . find a way of making judgements about new curriculum developments that is based upon exploration, experimentation, and measurement."

(36) One of the instruments which they used attempted to measure mental skills and attitudes. The categories which were scored, as reported by them, are given as follows:

Mental Skills:

(a) Analysis and interpretation of data

(b) Extrapolation

(c) Hypothesis formation

(d) Ability to design experiments

Attitudes

(a) Intellectual honesty

(b) Open-mindedness

(c) Critical mindedness

This test consisted of four questions. Each question contained a problem focus in the form of a paragraph followed by four alternative explanations. The students were required to discuss each explanation, select the most likely one, and describe a method which they would use to test the truth of the explanation. Marking procedure consisted of assigning values of 1 for clear evidence of a skill and attitude, ½ where the response failed to fully satisfy the criteria and a value of 0 for responses which showed little or no evidence of the skill or attitude.

Tests for marker reliability were applied separately to the skills totals and the attitude totals. Reliabilities ranged from .76 to .96 on the skills totals and from .48 to .66 on the attitude totals. The reliability of the test, using the Spearman-Brown prophecy formula, was calculated to be .89.

The attitude component measured in this test is closely associated with the intellectual behaviors of scientists as listed in An Inventory of the Affective Attributes of Scientists by Nay and Crocker (81) and is restricted to inferences made on the basis of the intellectual behavior exhibited by the student. In other words, the attitude measured by this test is closely allied to intellectual performance. This component is probably present in the other instruments cited although no provision is made to measure it separately. A second interesting feature of the instrument is the scoring procedure in which grades of competency are recognized. This procedure is a realistic one but it would be quite dependent on the consistency of the markers.

D. Summary

The tests, as reviewed above, lie mainly in the cognitive and psychomotor domain but some of them are also intimately associated with the affective domain, i.e. the tests which ask for responses about science and scientists could evoke responses associated with attitude and the Flynn and Munroe procedures attempt to measure attitudes related to the intellectual behavior of scientists.

Although no general all-purpose process test exists, some generalizations may be made. Kruglak (63) has emphasized and demonstrated that reliable tests for very specific skills may be constructed. The AAAS Science Process Instrument and the Burmester test of The Ability to Think Scientifically tend to substantiate Kruglak's conclusion. The technique of releasing information at the examinee's request is a useful technique in determining the efficiency in obtaining and using relevant information in order to derive a logical inference. Hubbard (53), Jones (57), and Suchman (114) used this technique but there is considerable latitude in the scoring procedures. Problem foci should be realistically portrayed with the emphasis on clarity of presentation. Diagrams may be superior to pictures for some presentations because the extra detail may mask significant elements and verbal descriptions may be adequate for some illustrations given to people with a relatively high degree of literacy. Because of individual differences, there may be some merit in varying the medium for the problem foci used for junior high school or elementary students. In this manner students of high verbal ability would have no special advantage if film loops were used and students with acute visual perception, who might have some advantage with the film loop presentation, would not necessarily enjoy this advantage with verbal input.

In any attempt at curriculum evaluation, the choice of the instruments depends on the purpose of the evaluation. Although there are a number of instruments available which purport to measure process type objectives, an investigator can do little more than proceed in a manner as outlined by Smith and Tyler (107); i.e., begin with the overall objectives and analyze them into behaviors which are testable prior to selecting, adapting, or preparing instruments to be used in the evaluation.

III. TECHNIQUES IN CLASSROOM OBSERVATIONS

In any curriculum evaluation in which there is concern for obtaining information pertinent to the improvement or development of an experimental curriculum, it is essential for the evaluator to determine the nature of the actual curriculum as distinct from the official curriculum. Comparative studies in which an experimental curriculum has been compared to a "traditional curriculum" have been barren of significant results other than expected results such as: BCCS classes outperform "traditional" biology classes on BCCS comprehensive tests (46), CHEMS students and students in "traditional" chemistry classes do not differ significantly in chemistry achievement as measured by "traditional" tests (97) and so on. There have also been some unexpected results. Mokosch (76) discovered that "traditional" classes outperformed process oriented classes on tests designed to measure competency in the process dimension. Taylor-Pearce (119) reported that students who were given an opportunity for creativity in a so-called "mathematizing method" were inferior to students in "traditional" classes as indicated by measures designed to measure creativity. When Crumb (28) allowed for differences in background among teachers, he found that PSSC students showed a greater understanding of science than did non-PSSC students but Trent (121) who did not allow for teacher variability found no significant difference in understanding science between PSSC and non-PSSC students. The effect of the teacher upon the official curriculum has been demonstrated by Gallagher. Gallagher (43) used a sample of six teachers of similar background teaching BSCS biology and reported from the observation and subsequent analysis of recordings of each class on three consecutive days that there was a wide variation in the topics dealt with and the manner in which the topics were handled. In effect each teacher presented his own curriculum so that any assumption of a common curriculum would be unjustified.

The problem of measuring effective teaching behavior is a difficult one and several approaches have been used. Some approaches are built on the assumptions that the viewer can recognize good teaching by simple observation. This is a fallacious assumption on two counts: first, the period under investigation and the interaction between teacher and student includes the effect of the term prior to the observation plus the prior activities of the teacher in what Jackson (56) calls "preactive" behavior; second, the observer has a preconceived notion of what good teaching is and this _a priori_ set can only lead to observations within the bounds of this set. Medley and Mitzel (71) report that ratings of this type show almost no discernible relationship to pupil learning.

Gage (38) suggests that it is possible to deal with specific aspects of teaching after knowing which teachers are effective and which are less effective. The behavior of these teachers would be studied and a video-tape recording made. This would allow for a multitude of re-runs which would be invaluable for the purpose of intensive observations. He suggests that a record of this behavior would be taken " . . . under unobtrusive conditions at the moment when the pupil learning was being engendered." Just how this could be done unobtrusively is not made clear but obviously a gross invasion of privacy would result if these recordings were made in a truly unobtrusive manner. Developers of observational systems in a democratic society must always be subjected to the obvious limitation that the system being observed is probably being affected by the intrusion of the instrument.

Reviewers, such as Gage and Unruh (39), Biddle and Adams (10), and Medley and Mitzel (70), have demonstrated that by 1968 there were more than thirty different classroom observational systems reported. In general, the reported systems were examples of empirical research attempting to relate measured aspects of teaching behavior to some criterion of teacher competency such as improvements in student achievement or behavior and ratings of the teacher by pupils or supervisors. Gage and Unruh (39) divide approaches to research on teaching into two parts. One approach deals with describing the way teaching is in the sense exemplified by Jackson (56) and the other deals with developing a model of instruction and fitting the teacher (or teaching device) to the model. This study is concerned only with systems which give some promise of validity and reliability in the actual description of classroom behavior of teachers.

Most of the research systems utilize classroom verbal behavior as the main source from which to draw inference about the classroom behavior of teachers. Smith (109) may be overemphasizing the importance of verbal behavior when he states: "Teaching is almost entirely verbal behavior . . . ." but there is little doubt that verbal activity forms part of most classroom activity. The analysis of verbal behavior thus remains an important part of most observational systems.

Most observational systems utilize a coding system in which an observer records his observations into a form which may readily be tabulated. Flanders (35) considers that the most important criterion for any coding system is that a trained person can reconstruct the recorded behavior from encoded record even although he was not present at the observation. This is basic to methods developed by Flanders

(86) and by Flanders and Amidon (86).

Flanders Interaction Analysis Technique (35) requires an observer in the classroom to make judgments and corresponding categorizations approximately every three seconds. Given n categories, arabic numbers from 1 to n are used to code the categories. The observer simply writes down the number corresponding to the category of the observed event in such a way as to preserve the original sequence. The data is transferred to an n by n matrix for purposes of analysis. This procedure allows for a relatively large number of entries (approximately 800 during a 40 minute period) and thus allows for a large number of elements in the sample of classroom behavior. The short time interval allowed for each categorization practically necessitates the use of trained and practised observers. Flander's technique has been widely used and Medley and Mitzel (71) report that systems using recorded observations at regular time intervals for pre-selected categories have succeeded in identifying aspects of the affective climate of the classroom.

A number of researchers such as Smith and others (109) and Nuthall (85) have adopted units of classroom observation such that each unit corresponds to an analytic definition corresponding to the conceptual assumptions of the investigator. The use of analytic units generally requires some permanent recording such as sound or audio-visual recording. Problems of semantics make systems using analytical units difficult to translate into a form readily understood by another investigator of different conceptual posture.

Medley and Mitzel (71) indicate that measurements of classroom behavior are subject to the bias of the observer and, in many cases, uninterpretable by anyone. In an attempt to increase the scope and reliability of classroom observation procedures, they developed the Observation Schedule and Record (OScAR). The OScAR is a technique for objectively observing and recording classroom behaviors. The OScAR provides a schedule for recording behaviors and assists the observer by also providing suitable cues. Data on activities, grouping, materials, and symptoms of classroom climate are recorded during a five-minute period. Following the first five-minute period the subject area is recorded and a second five-minute period is devoted to tallying indications of "expressive behavior" in each statement made by the teacher. Following the recording of the subject area, a third five-minute period is devoted to the recording of activities, etc. This alternation of focus of interest is continued for six five-minute periods.

Medley and Mitzel (72) collected data in 49 classrooms and after reliability considerations combined the items into 14 "keys" which were factor analyzed and three orthogonal factors identified. These results indicated to them that the OScAR gives reliable information about:

1. The social emotional climate
2. The relative emphasis on verbal learning
3. The degree to which the social structure centers about the teacher

Thus the OScAR does not appear to be getting at much more than measures of "affective climate" comparable to measures using the Flanders' system. Medley and Hill (74) have indicated the similarity of the results obtained using the two systems. They intercorrelated 75 measures (38 from the Flanders' record and 37 from the OScAR record) taken from two parallel records from each of 70 teachers. A factor analysis was performed on the correlation matrix. Ten factors were identified of which five were measured by both systems, three by the OScAR system, only, and two by the Flanders' system only. There would appear to be a great deal of overlap between the two systems but the OScAR is more suitable for use by relatively untrained observers. Medley and Mitzel (73) found no relationship between the information measured by the OScAR and the growth of elementary students in reading ability or in group roblem solving skill. There is thus little evidence that the OScAR gets at aspects of behavior related to cognitive objectives. It is conceivable that the OScAR could be expanded to include additional cognitive categories.

Gallagher (43) has developed an observational system referred to as the "Topic Classification System" (TCS) which includes cognitive variables. The Topic Classification System was developed as a result of the experience with a previous system developed by Aeschner, Gallagher, and others (43). The previous sytem was based upon the structure of intellect model of Guildford (48) and contained four categories corresponding to Guilford's five cognitive categories and a fifth category for miscellaneous items. The categories of cognitive memory, convergent thinking, divergent thinking, and evaluative thinking correspond to the cognitive processes of memory, cognition, divergent and convergent production, and evaluation. The Topic Classification System attempts to indicate three dimensions of classroom behavior, namely, the level of instructional intent, the level of conceptualization, and the level of style. The style dimension appears to correspond to the four-category system used by Aschner, Gallagher and others (43), but the levels of style are specific in nature and thus more readily identified than categories corresponding to Guildford's classification. With the three dimensions including five divisions of style, two of instructional intent, and three of conceptualization, there are thirty cells potentially available for analysis.

The potential range of the Topic Classification System is reminiscent of the work of Siegel and Siegel (107) who used a multivariate approach for studying interaction among variables. The approach of Siegel and Siegel involves four classes of independent variables (learning environments, instructors, learners, and courses) and the dependent variables of effectiveness of achievement, process of achievement, attitude, thought, and out-of-class behavior. Because of the large number of variables included, the identification of specifics would be relatively difficult in the subsequent analysis. Gallagher's TCS has already chosen specifics which he calls "levels" and if the coding procedures are adequate, differences among teachers in terms of these specifics can readily be obtained.

The TCS with its three dimensional character has the potential for providing a great deal of information about classroom behavior. Probably its most serious disadvantage, from this researcher's point of view, lies in the fact that the identification of each topic requires the use of trained individuals working with transcribed verbal data. It would seem that certain variations, including timing procedures associated with the identification of non-verbal learning activities, would have to be included in the system before it could be suitable for identifying or classifying "significant incidents or units" in classrooms in which non-verbal activities form an important part of the instructional process.

Fischler and Zimmer (34) have developed an observational instrument which is specifically designed for use in science classrooms. As reported by the authors:

> In common with most time-sampling instruments, it had the following features: (1) observation was made by an eye witness, (2) the behavior to be observed was defined in terms of overt action, (3) the behavior of an individual or group was observed for a stated unit of time; usually short, (4) there was a number of repetitions of the time unit employed, (5) an individual score was based upon the number of time units in which the defined behavior occurs. (32)

The instrument contains explicitly described categories listed under three main headings: (1) Teaching Techniques,

(2) Teacher's Questions, and (3) Characteristics of Teaching. The observed behaviors are checked off every two and a half minutes and provision in made for recording what students and teachers do as well as what they say. The instrument was field tested and as a result of that experience the categories and definitions were revised and refined. Although Fischler and Zimmer claim that the instrument enabled them to, " . . . differentiate the teaching characteristics of individuals teaching science.", they do not document this claim with empirical data.

The Observational Instrument for Science Teaching (32) as described above, utilizes a much longer time interval than the Flanders' system and thus decreases the number of entries but this makes it feasible for the observer to make judgmental decisions covering a relatively large area of concern. The categories as defined by the instrument correspond to activities specific to science classrooms and include a selection of the "instructional modes" as outlined by Powley (94).

It seems apparent from the observational systems considered here and from information obtained from the reviews of Medley and Mitzel (71), Gage and Unruh (39), and Biddle (10) that observational procedures which are based upon time sequences and which use explicitly described categories are more amenable to the use of relatively untrained observers than are systems based on analytic or phenomenal units.

Category-time sequence methods have two main concerns. One is associated with the evaluative judgment of the rater and the other with obtaining a representative sample of the behaviors being observed. The first may be partly alleviated by the use of very explicit instructions and by limiting categories, where possible, to specifics. The chance of obtaining a representative sample would be enhanced by making sure that the observers had sufficient time to reliably categorize the observed behaviors with the time unit as short as possible within these bounds.

The use of structured observations is one of the most defensible methods of describing classroom behavior. Grobman points out that with appropriate techniques and their subsequent analysis:

> It is possible to describe verbal and non-verbal classroom behavior in relatively precise terms. It is also possible to determine whether the interaction pattern in the classroom is compatible with the project objectives. (47)

Stake considers such descriptions to have a high priority in curriculum evaluation when he states:

Neither an understanding of what the curriculum has been or what should be tried next time is possible without data on the teaching methods. In some evaluation studies the most valuable data will be those gathered by a classroom observation system. (110)

CHAPTER III

THE DESIGN AND DEVELOPMENT OF EVALUATIVE PROCEDURES

I. THE EVALUATION MODEL

This writer became associated with the Edmonton Junior High School Process-Approach Science Project approximately two years after its formation and became interested in the evaluative aspects of the program. This evaluation was conceived to be evaluation of student achievement in the process dimension in relation to teaching performance and curriculum materials in an attempt to determine what sort of instructional procedures produced particular kinds of achievement in students. At the time this study was being planned, the use of the Inventory was extended to biological science classes at the grade seven level. It is to the curriculum thus established that the present study is directed.

Although this study was not directly concerned with the affective domain, it was recognized that students with negative attitudes toward school tend to be less successful in school achievement than those with positive attitudes. The close relationship between attitude and achievement has been recognized by Maykovitch (69) who found that the attitudes of students entering high school was closely related to school achievement but subsequent changes in attitude did not appear to produce any great change in achievement. Aiken (1) reports that for students at the high school or college level: " . . . attitude scores contribute something over and above ability test scores, to the prediction of achievement in mathematics." Because students do enter junior high school with attitudes which could be relatively stable over a period of time, it was considered essential to take cognizance of a possible attitude effect in an attempt to minimize its effect as a moderator variable.

The students of all teachers taking part in the experimental curriculum formed the sample. This sample was further broken down into treatment groups of one group for each teacher or method. Each method was considered to be distinct, having in common only the curriculum materials and the structure of these materials within the framework of the Inventory. In this manner all treatment groups were subject to the same overall behavioral objectives but not to the same methodology. It was therefore considered necessary to describe each method using procedures aimed at determining what actually took place, both from a long range view and from a more intensive direct view.

Adopting the terms used by Jackson (56), it was considered essential to obtain information about the preactive and interactive behavior of each teacher. (Interactive behavior refers to behavior in the classroom in which the teacher interacts with the students and preactive behavior refers to behaviors which are relevant to teaching but do not include interaction with the students.) Following the evaluation of student performance upon selected evaluation instruments, the information obtained about each method was related to the evaluation of student achievement. The diagram below illustrates an instructional model which shows the relationships between the main concerns of the study.

In this study, in terms of the above model, information was gathered relative to the preactive and interactive behavior of the teachers and, at the same time, part of the instructional model was formed by carrying out an evaluation of student performance relative to the behavioral objectives of the experimental curriculum.

The behavioral objectives were simply those as implied by the science processes as set out in the Inventory.

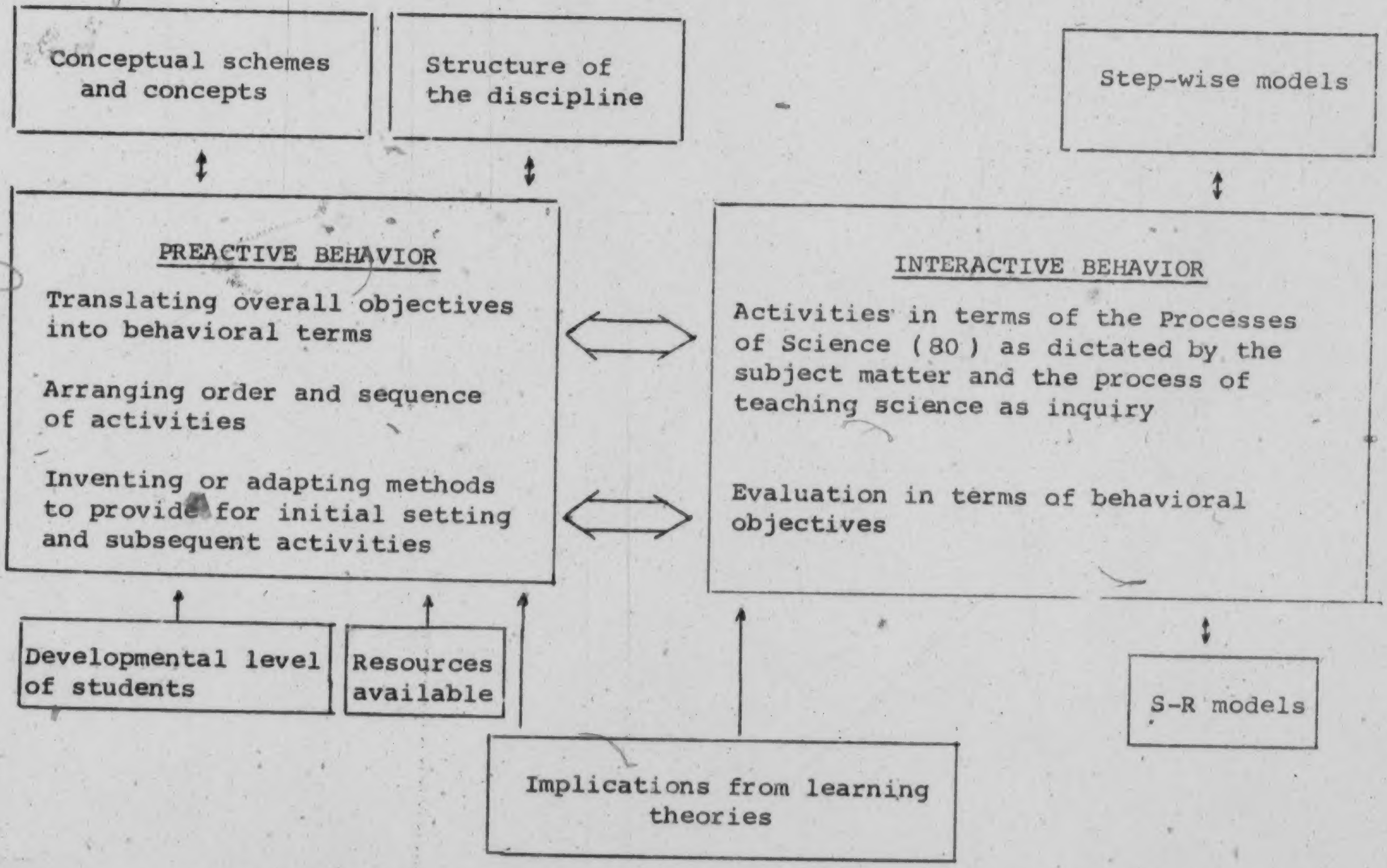

Conceptual schemes and concepts
Structure of the discipline
Step-wise models
PREACTIVE BEHAVIOR
Translating overall objectives into behavioral terms
Arranging order and sequence of activities
Inventing or adapting methods to provide for initial setting and subsequent activities
INTERACTIVE BEHAVIOR
Activities in terms of the Processes of Science (80) as dictated by the subject matter and the process of teaching science as inquiry
Evaluation in terms of behavioral objectives
Developmental level of students
Resources available
Implications from learning theories
S-R models

Figure 1. An instructional model

That is, each process when prefixed by the words: "The student should be able to" and altered slightly in the interests of grammatical correctness was considered to be a behavioral objective. These behavioral objectives were common to all groups by agreement with the participating teachers; i.e., it was established at an in-service session prior to any pretesting of students that teaching within the framework of the Inventory meant the acceptance of the aforementioned overall behavioral objectives.

It was considered reasonable to believe that individuals become more proficient in what they practice. Teachers participating in the experimental curriculum had accepted the behavioral objectives as implied by the Inventory and the students of these teachers would (or should) be using the processes of scientific inquiry in dealing with "investigations." The students of these teachers should therefore be gaining skill in, and knowledge of, the processes of scientific inquiry as set out in the Inventory. Students who actively participated in activities which required that they deal with scientific investigations should become more efficient in knowing how to deal with a real or simulated scientific investigation. The main instructional strategy of using the processes of scientific inquiry in conjunction with investigations is one in which the student "imitates" the scientist and thus gains familiarity with the tactics employed by scientists.

The instruments selected, prepared, or adapted for the evaluation of student achievement of behavioral objectives were expected to probe the students', (a) knowledge of, and skill in, the processes of science, (b) knowledge about scientists and the scientific enterprise, and (c) efficiency in inquiry procedures. Inquiry procedures in the sense used in (c) refers to inquiry in which according to Suchman (114) the student (1) searches (2) processes data (3) discovers, and (4) verifies. Essentially the evaluation procedures were designed to harmonize with the overall objectives of the experimental curriculum.

Although the main instructional strategy of the experimental curriculum was to have the student "imitate" the scientist, it was also recognized that there is a place for other instructional strategies in specific instances. Step-wise or hierarchal models may be used quite effectively for the development of specific knowledges or skills and rote learning by means of S-R models may be desirable in some instances. There is probably merit in having students of biological science deal with simple exercises in classification and continue in hierarchal fashion through more complex examples. Similarly, chemistry students could become familiar with the symbols of the more common elements by means of an effective use of S-R models.

This researcher had little control over the left-hand side of the foregoing model, (Figure 1). The structure of the discipline and the conceptual schemes and concepts which influenced the teachers' preactive behavior were largely determined by the fact that all of the students in the sample were studying biological science at the grade seven level and had access to the same textual and reference materials. (Conceptual schemes and concepts are not included as part of the discipline in the recognition that although each discipline has its own set of conceptual schemes and no one set can apply to all disciplines, there are some conceptual schemes common to more than one discipline.) The overall objectives, as explained above, were the same for all groups but each teacher was free to select and adapt methods which he considered to be most suitable for the attaining of these objectives.

This researcher rejected the idea that, when classroom instruction is considered as a treatment, treatment could be specified in advance. Instead he adopted the point of view that the nature of each treatment had to be estimated from information related to what actually took place within the classroom.

Essentially this study resolved itself into two interdependent parts as illustrated in Figure 2.

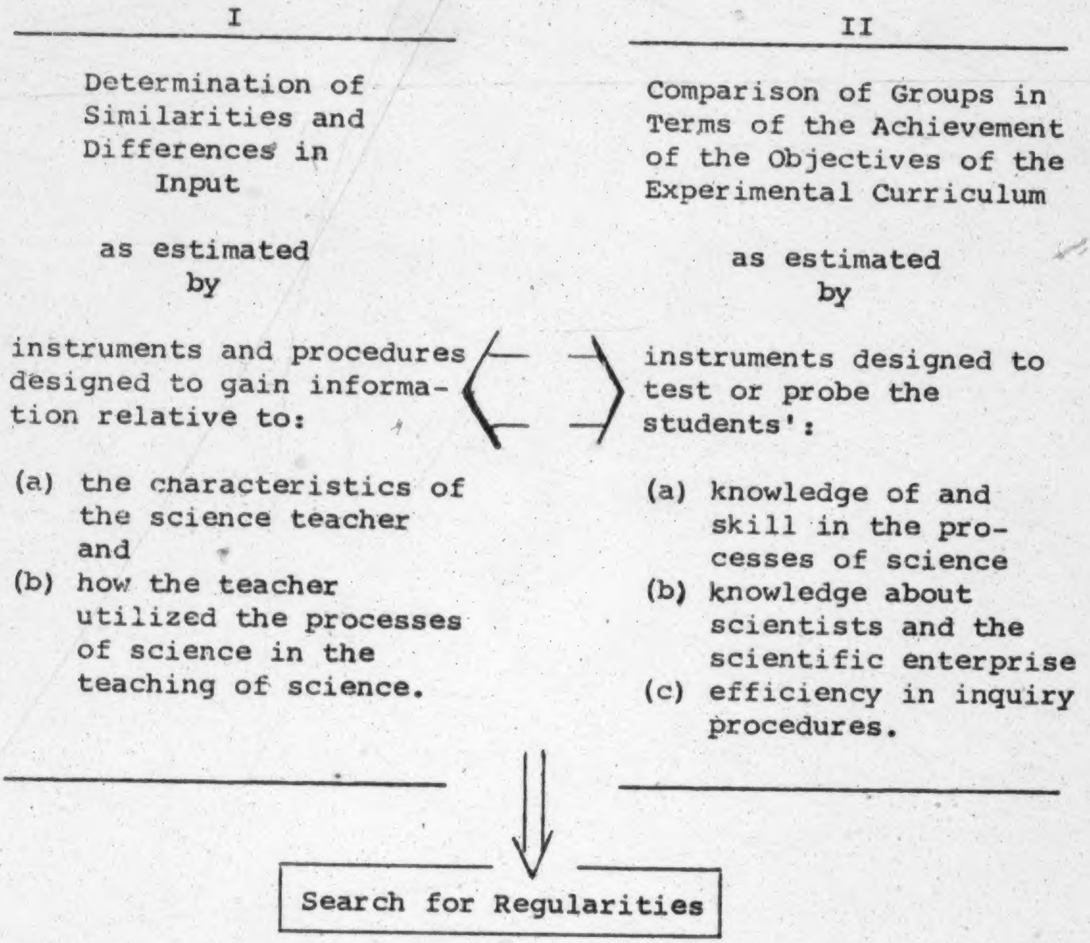

I
Determination of Similarities and Differences in Input
as estimated by
instruments and procedures designed to gain information relative to:
(a) the characteristics of the science teacher and
(b) how the teacher utilized the processes of science in the teaching of science.
II
Comparison of Groups in Terms of the Achievement of the Objectives of the Experimental Curriculum
as estimated by
instruments designed to test or probe the students':
(a) knowledge of and skill in the processes of science
(b) knowledge about scientists and the scientific enterprise
(c) efficiency in inquiry procedures.
Search for Regularities

Figure 2. The evaluation model

It was anticipated that the information obtained about the similarities and differences in input could be matched to the information obtained from the comparison of the achievement of groups to reveal some regularities. If such regularities existed, some inferences could conceivably be drawn about the kind of input which tended to maximize process approach objectives.

II. THE SAMPLE

All students of teachers using the experimental program formed the sample provided these students remained in attendance in the same class for the complete period under observation. The sample was thus composed of sixteen classes or a total of 373 students. The teacher participants were volunteers who agreed to, (a) teach within the framework of the Inventory, (b) submit a daily log book of their teaching activities with their experimental classes, (c) allow a team of observers into their classes for a minimum of three sessions, and (d) participate in a series of in-service sessions.

The teacher participants could not be considered to be representative of the population of junior high school science teachers in the city of Edmonton. In the view of this writer, these teachers were superior teachers who were quite interested in improving the quality of science education. All of the teachers had at least two years experience with the *Inventory* and four of the five active participants were junior high school science coordinators in the school system. (Each coordinator had responsibility for several schools. He acted in an advisory capacity and attempted to coordinate science activities among the schools under his surveillance.)

The in-service sessions were held every third week beginning with the initial meeting during the second week of September. A total of six were held in all. The sessions were chaired by David Powley (94) who was interested in developing instructional techniques suitable for integrating the process dimension into a science curriculum. Part of the first session was devoted to the establishment of agreement among the participants about the amount of time available to this researcher and his co-worker such that the researchers would obtain the required information with minimal disturbance of classroom routine. The remainder of the first session consisted of a rather free-wheeling discussion of the experimental curriculum in which the teacher participants exchanged views on methods of teaching within the framework of the *Inventory*. During the second session this researcher outlined his plans for the evaluation aspects of the experimental curriculum. The remaining four sessions were devoted to consideration of various instructional modes suitable for use in the process dimension. The explanation of the possible uses of these modes was carried out by Powley (94). It is to be clearly understood that these sessions were considered to be seminars participated in by equals; and although Powley presented demonstrations of possible methods for integrating the process dimension into a science curriculum the teachers were under no obligation to accept any suggestions.

Although the teachers appeared to form a somewhat homogeneous group, some differences were suggested during the discussions which took place during the first two in-service sessions. This researcher made notes about the spontaneous debates which occurred during the first two sessions and then listened to a taped recording of these sessions. It seemed evident that the teachers, designated hereinafter as A, B, C, D, and E, respectively, emphasized approaches to the junior high school curriculum such that each approach had an emphasis characteristic of the indivi ual teacher. The impressions gained somewhat intuitively by this researcher suggested the following characteristic emphases:

Teacher A - Loosely structured inquiry, emphasis on discovery

Teacher B - Emphasis on processes as listed in the *Inventory* and occasionally studying the processes in isolation

Teacher C - Subject matter leading rather than lagging the inquiry.

Teacher D - Processes of science not emphasized by name; assimilation by utilization emphasized

Teacher E - Utilization of processes. Naming of specific processes considered to be important.

The sample was considered to be composed of six treatment groups of grade seven students, designated hereinafter as groups A, B, C, D, E, and X, respectively. Groups A to E inclusive were comprised of students of teachers who actively participated in all phases of the program, whereas group X was comprised of two classes of one teacher who had not volunteered for the program but did teach within the framework of the *Inventory* and had cooperated closely with one of the active participants. This group (X) was included because groups E and X met as a large group for some large group instruction.

Table I shows the distribution of the students by

TABLE I

DISTRIBUTION OF STUDENTS

GROUP		A			B			C	C		D	D	E	E	X	X
CLASS	1	2	3	4	5	6	77	8	9	10	11	12	13	14	15	16
Boys N	11	14	14	9	6	14	16	14	13	17	12	12	9	10	10	18
Girls N	15	16	12	15	14	3	11	8	8	-	15	9	7	13	16	8
Class N	26	30	26	24	24	17	27	22	21	17	27	21	16	23	26	26
Group N		82			65			87			48		39		52	
Total N					373				Boys N 199					Girls N 174		

sex, class, and experimental group. Five schools were involved in the study as groups D and X came from the same school and shared the same facilities. Although no information was available to indicate the economic status of the parents of the students, it was noted that groups B, E, and X attended schools in newly developed areas of the city whereas groups A, C, and D attended schools in established areas. For the three groups in established areas, the economic status of the neighborhood would place group D on the highest level economically and group A on the lowest.

III. TESTING PROCEDURES AND TREATMENTS

An important part of this study as illustrated in the evaluation model (Figure 2) was the assessment of student achievement. It was realized that excessive testing could cause students to become resentful or bored and thus introduce an upsetting factor which is difficult to control. This suggested that a balance must be attained between student tolerance of a testing program and the obtaining of adequate information. Crucial to this aspect of this study was the attitude of the participating teachers. Their full cooperation was assured providing all student evaluation was confined to regular class periods and limited ~~to a~~

to a maximum of seven periods. This arrangement allowed for all testing which was considered to be essential.

Treatments were similar only in the sense that each teacher operated within the framework of the same overall objectives as implied by the Inventory. This assured (barring undue outside influences) a set of treatments whose differences would be effected by each teacher's interpretation of the overall curriculum and the methods and techniques employed by each to attain the objectives of this curriculum. The nature of the differences among treatments was determined from daily log books submitted by each teacher and from classroom observations made in accordance with pre-planned observational techniques using an instrument prepared for this purpose.*

The groups were pretested during the first two weeks of October using the Test On Understanding Science Form eW (TOUS) and the Cooperative General Science Test-Form B. The tests' scores relating to I.Q. were available at the schools as all students who had been in grade six the previous year had written the Lorge-Thorndike Intelligence Test (verbal and non-verbal). During the month of February five tests were administered. These included TOUS as a posttest, the *Science Reasoning Test* (SRT), the *Process of Science Test* (PST), the *Inquiry Efficiency Test* (IET), and the *How I Feel About My School* (HIFAMS) attitude test.

---

*Copies of the Log Book and Observational Instrument are presented in Appendix B.

In terms of the evaluation model (Figure 2), the TOUS was used to measure knowledge about scientists and the scientific enterprise. The *Process of Science Test* (PST) and the *Science Reasoning Test* (SRT) were designed to measure knowledge of and skill in the processes of science. The *Inquiry Efficiency Test* (IET) (a modification of the *TAB Science Test* (57)) was used to compare students in terms of their ability to select questions whose answers are relevant to the solution of an unfamiliar problem in science. The latter three instruments were prepared or modified for this study and, although field tested in a pilot study, are relatively untried instruments which should be interpreted with some caution.

The aforementioned three science tests depended upon student reaction to a new or novel situation hence a repetition of them would have little meaning. It was considered impracticable to construct a parallel form of each using the criteria suggested by Gulliksen (49) because each test involved several "investigations" which, due to the small number per test, could not be considered as a sample of the domain of possible investigations. This would make the probability of equal means and standard deviations somewhat unlikely. Even if parallel forms were available, the limited testing period would necessitate the elimination of other instruments considered to be essential for this study. From these considerations it was decided to use the PST, IET, and SRT as posttests only. A description of the instruments and evaluative procedures follows.

A. Tests and Instruments

Co-operative Science Test. This instrument has been widely used and statistics relating to it are well documented (25). It consists of sixty multiple choice items and has a time limit of forty minutes. The statistics relating to it were taken from populations which may not be representative of the sample used in this study. The scores obtained from the administration of this instrument in the present study were analyzed and the reliability calculated using the Kuder-Richardson Formula 20. The KR-20 reliability coefficient calculated for the sample of 373 students equalled .86. This compares favorably with the reliability estimate of .90 published in the Co-operative Science Test Handbook. (25).

Test on Understanding Science (TOUS).* This test, as indicated previously, has undergone considerable testing and has been widely used in studies dealing with curriculum evaluation. The TOUS tests, of which the form Ew is suitable for junior high school students, are designed to sample three areas of understanding: understanding about the scientific enterprise, understanding about the personal characteristics of scientists, and understanding about the methods and aims of science.

Using the Kuder-Richardson Formula 20, a coefficient of reliability of 0.64 was obtained from the pretest data and a coefficient of reliability of 0.72 on the posttest data.

The items in the TOUS test were grouped into three main categories according to the area of understanding being sampled. This was accomplished with the assistance of four judges who were specialists in science education. Each of the judges was given a copy of the test and was asked to list the items which seemed to belong to each of the three areas of understanding listed previously, namely: (a) understanding about the scientific enterprise, (b) understanding about the personal characteristics of scientists, and (c) understanding about the methods and aims of science. Although the classification was done independently by the judges, agreement was achieved on thirty-two of the thirty-six items. Unanimity of agreement was achieved by consultation with the judge who failed to agree with the majority. In all four instances the disagreements were due to a failure on the part of some of the judges to consider the complete item being judged rather than some one particular part in the stem or the body of the item.

---

*See Appendix C p.273 for a copy of this test.

Items 2, 5, 9, 10, 13, 22, 24, 29, and 33 were allocated to (a) as categorized in the previous paragraph, items 6, 8, 12, 18, 20, 26, 31, and 36 to (b), and items 1, 3, 4, 7, 11, 14, 15, 16, 17, 19, 21, 23, 25, 27, 28, 30, 32, 34, and 35 to (c). In this manner three sub-tests were formed which will subsequently be referred to as TOUS(a), TOUS(b) and TOUS(c) respectively.

Inquiry Efficiency Test (IET).* This test is simply an adaptation of the TAB Science Test (57) and consists of a combination of Form A and Form B. This procedure was necessary because one film from each form had already been viewed by some of the classes in the sample.

---

*See Appendix C p.284 for a copy of this test

As adapted, this test consists of three sections. In section 1 a physics problem focus film is presented and the student is asked to select the most correct explanation from a given set. In section 2 the student is presented with a list of questions answerable by a "yes" or "no" response. The correct response ("yes" or "no") follows each question but is completely masked by easily erasable material. The student gathers clues one at a time by erasing the material covering the correct answer. Finally in section 3, the student is presented with the same set of explanations as those found in section 1. When he is confident that he knows the correct answer (or when he has exhausted all of the questions of section 2) he may select the explanation he considers to be correct by erasing the material covering the "yes" or "no" opposite each explanation. A "yes" indicates that his selection is the correct one. Should he fail to get the correct answer on the first try he is entitled to go back to section 1 to search for more information.

Scoring for the test is based upon assigning a score of 20 for the correct solution to each problem. This score is reduced by 1 point for each question "asked" in section 2 with the proviso that one question may be "asked" without penalty. A failure to select the correct explanation from section 3 is assigned a penalty of 6 points for each wrong selection.

This method of scoring was chosen because of its relative simplicity and because it promised to avoid procuring data containing apparent discontinuites. This would make the data more amenable to subsequent analysis. The child was thus not penalized for not gathering information. This writer is inclined to the view that not enough is known about whether an "optimal sequence" does in fact exist, to penalize a student for failing to select a predetermined one. The gaps in knowledge which exist in children depend on so many factors that it was considered unwise to arbitrarily assign weights to individual questions.

Science Reasoning Test (SRT).* This test is entirely verbal and has a format similar to the test mentioned in Chapter 2 developed for college students by

---

*See Appendix C p.296 for a copy of this test

Mary Alice Burmester (17). Initially seventy items were prepared and designed to match the processes of science as set out in the Inventory without placing undue emphasis upon the name of each particular process. A deliberate attempt was made to use synonyms and associated description to avoid giving advantages to students who had been drilled on the names of the processes of science as listed in the Inventory. These seventy items were submitted to a group of ten grade seven students in an informal setting under the supervision of the researcher. The students answered the questions with the understanding that they were free to ask questions about anything which they did not understand. This preliminary work was quite helpful because of the spontaneous nature of the students' comments and reactions and provided advance notice dealing with readability level, time requirements, and the difficulty level of the questions. In view of these gross indications, revisions were made and the test adjusted to contain 40 items.

The test in this form was submitted to a panel of four judges who met in committee with the author. An item was considered to be satisfactory if it matched one of the seventeen processes of the Inventory and was free from ambiguities. Items were revised in accordance with suggestions which formed the consensus of the committee. The revised form was then used in a small pilot study as detailed below. Minor changes in wording were made and the test was reduced to contain 35 items in order to give ample time for the test to be completed during a normal classroom period.

The coefficient of reliability was estimated from the data obtained in the pilot project. The Kuder-Richardson Formula 20 gave a value of .78.

Process of Science Test (PST).* For this test two films were selected as simulated investigations in science. These two films were selected using the criteria that each represent a scientific investigation in Life Science whose subject matter would be unfamiliar to the majority of grade seven students. The two films which met these criteria were BSCS Inquiry Films, one entitled Phototropism and the other entitled The Amoeba. It was necessary to block out the titled parts of the film by using a small card which enabled the administrator of the test to view the film although it was blocked off to the class.

This required considerable practice and made it necessary for the administrator of the test to engage in a series of prior viewings in order to know precisely when the sub-titles were to appear.

---

*See Appendix C p.304 for a copy of this test

Questions based on the Inventory were prepared and submitted to a group of ten grade seven students in an informal setting as was done for the Science Reasoning Test. After making revisions which seemed desirable, copies of the test were given to each of four judges for their comments and evaluations. Minor changes in wording and format were made and the test was used in a small pilot study in conjunction with the Science Reasoning Test. As a result of the experience in the pilot study, further revisions of a minor nature were made. In its final form the test was made up of two main investigations with eleven items associated with each investigation. This form was used in the major project and the K-R20 reliability calculated. This gave a reliability estimate of .53. As the K-R20 reliability is essentially a measure of internal consistency, this relatively low value probably reflects the diversity among the items.

The Pilot Study. The two instruments which were specifically designed to measure the process dimension were the Science Reasoning Test and the Process of Science Test.

It was considered essential to test whether these instruments gave some promise of validity before using them in the project.

Two classes of grade seven students were selected for this purpose. The classes were approximately equal in terms of I.Q. scores on file at the school. One of these classes was taught by a teacher who was quite familiar with the purpose and use of the Inventory. The other class was taught by a teacher who was not familiar with the Inventory or its purpose. The Science Reasoning Test was administered to each class on consecutive periods of the same day and a similar procedure was followed for the Process of Science Test one week later. For purposes of identification, these classes are classified as process-oriented and traditional respectively.

The results of the pilot study are presented in Table II. In both cases the process-oriented class did better. In the comparison of the PST scores, the difference between classes was significant ($p = .03$). When it was considered that this test used visual presentations as problem foci, it was believed to more closely approach the reality of a practical scientific problem involving laboratory investigations. Students who operate well in the

TABLE II

PILOT STUDY DATA FOR PROCESS OF SCIENCE TEST AND SCIENCE REASONING TEST

	Group I Process-oriented Class N = 28		Group II Traditional Class N = 28		
	X	S.D.	X	S.D.	
I.Q. Verbal Lorge-Thorndike	105.5	13.2	104.8	13.9	
I.Q. Non-verbal Lorge-Thorndike	107.3	13.4	106.9	14.5	
Science Reasoning Test (SRT)	17.2	6.0	14.1	5.5	F = 3.36 p = .07
Process of Science Test (PST)	9.5	2.7	7.8	2.1	F = 5.76 p = .03

process dimension should therefore be expected to achieve better in this type of test than those who have not developed this ability. The more nearly a problem resembles a scientific investigation, the greater should be the relative spread between process-oriented classes and traditional classes. This last statement merely reiterates a truism that people tend to do better in tasks which are familiar to them and in which they have had an opportunity to participate and practice.

How I Feel About My School (HIFAMS).* This instrument was prepared by Coster and modified by Pyra (96). It consists of 31 items and purports to measure a student's attitude toward school from student responses recorded on a five point Likert type scale.

Observational Instrument.** This instrument was prepared and used jointly by this researcher and Powley (94). It consists of two parts on which observations are recorded concurrently. It utilizes a time sequence technique in which recordings are made at two minute intervals.

Part I of the Observational Instrument was designed to aid in validating the reports of long term procedures as

---

*See Appendix C p. 315 for a copy of this instrument

**See Appendix C p. 326 for a copy of this instrument recorded by the teachers in log books specifically designed for this purpose. This portion of the instrument attempted to determine what meaning could be attached to the processes of science as identified by the teacher. Part II of the Observational Instrument was an adaptation of the science observational instrument as prepared by Fischler and Zimmer (34). This portion of the instrument was used to gain information about student-teacher interaction and to obtain a profile of individual teaching behavior.

The instrument was field tested by two observers and a taped recording of the observed lesson was obtained simultaneously. By prior agreement the observers agreed to record the main categories by a diagonal stroke and to use a check mark for subsidiary categories noted during the interval. To avoid any tendency to limit observations to a localized section of the classroom, the classroom was divided into quadrants and the observers shifted their attention to the next quadrant in a counterclockwise direction at intervals of ten minutes. Immediately after the observational period of one class period duration, the observers checked for areas of disagreement and listened to a playback of the corresponding tape in order to improve the consistency of the results for subsequent observations.

Minor differences were noted in Part II of the Observational Instrument. These differences were due almost in their entirety to variations in the interpretations made by the respective observers. Agreement was arrived at in this area before the instrument was used in the project.

Part I of the Observational Instrument listed the seventeen major process categories and had associated with it a teacher report format which also listed these seventeen processes. Immediately following each lesson the teacher listed the processes with which he had been dealing. Part I of the Observational Instrument was designed to determine whether the teacher's conception of the particular science processes agreed with the criterion or standard established by the observers (judges).

The judges' standard was established by prior decision that any disagreement existing between judges would be resolved by means of the taped recordings of the classroom discourse. The time of the disagreement would be noted from the observational instrument and the time sequence could then be checked from a playback of the taped recordings. No categorization was to be considered valid unless both judges agreed upon a particular correction and this judgment was verified by consultation with two other independent judges familiar with this particular aspect of science education. In no case was it found necessary to resort to this procedure as there was perfect agreement between the observers on the major categories of Part I of the instrument. (The judgments for this part of the instrument required straightforward decisions such as: "Are they dealing with observations?". Unanimity of judgment here was not entirely unexpected.)

To obtain some measure of the reliability of the instrument, a chi-square test was performed on Part I of the instrument to test agreement between judges in the subcategories. The formula used in calculating the value of chi-square for each test was the familiar:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E)^2}{E}$$

In this formula, E is the expected and, O the observed frequency. Under the null hypothesis to be tested, the estimated frequency would be the mean frequency assigned to each category by the two observers. It was considered that if the null hypothesis could not be rejected for the subcategories, then reliability or consistency of judgment would practically be assured for the major categories. For Part II a chi-square test was also performed to test the consistency of the judgments made by the observers.

Categories in this section required judgments involving differences in degree, in some instances, so a rejection of the null hypothesis that no difference existed between judges would tend to be more probable than if only discrete categorizations were involved.

In Table III which follows, the critical values ($p < .05$) are given for not rejecting the null hypothesis plus the approximate probability of obtaining a chi-square less than the calculated value.

TABLE III

CHI-SQUARE CALCULATIONS TO TEST OBSERVER AGREEMENT

Observational Instrument	Degrees of Freedom	Chi-Square (calculated)	Critical Value ($P < .05$)	Probability of Obtaining a Chi-square Less Than Calculated Value
Part I	9	.72	16.92	.05
Part II	20	9.07	31.41	.05

The null hypothesis is not rejected for either Part I or Part II of the instrument, so on this basis the observers were not statistically different. This, in itself, is not necessarily a strong argument for reliability claims but when the low probability level for obtaining a lower chi-square than the calculated one is considered, some weight is added to the claim that the instrument, as used, produced reliable data in the inter-rater reliability sense.

B. VALIDATION PROCEDURES

In essence, validation deals with whether a test measures what it purports to measure. The procedures described herein deal mainly with attempts to determine what, in fact, was being measured. Four categories are commonly used to classify validation procedures. These categories, as classified by the American Psychological Association's Technical Recommendations are designated as content, concurrent, construct, and predictive validity. According to Thorndike and Hagen (121), there are two main types of evidence bearing on the validity of a test namely, rational and empirical. The rational evidence for this present study depended largely on the opinion of judges and the close relationship of the tasks required in the test to tasks required in actual classroom activities. These procedures have been described previously so the following will be concerned mainly with presenting evidence of an empirical or statistical variety.

Concurrent Validity of the Science Tests

A basic assumption was made that the teachers co-operating in the project were capable of judging student performance in carrying out scientific investigations. Accordingly each teacher was supplied with a key developed by Wilson (129) for the grading of laboratory performance in Biology courses (See Figure 3). Each teacher was given a prepared list of his students and instructed to assign a grade in accordance with the key. No further elaborations or instructions were given. The assigned grade was used as the criterion score and relationships between the criterion scores and each of the science test scores were obtained.

For purposes of analysis the Science Reasoning Test was divided into three parts. SRT 1 contained questions dealing with recognition of the phases of a scientific investigation, SRT 2 contained questions which require a student to exercise judgment about the suitability of an experiment for a given problem, and SRT 3 contains questions which require a student to evaluate facts related to substantiation of a given hypothesis or the refutation of this hypothesis. The Process of Science Test was divided into two parts. Each part contained questions related to a

Figure 3: Teachers' Guide For Assigning Student Grades

ESTIMATED GRADE FOR STUDENT
(KEY)

7	Excellent student. Does all work enthusiastically, conscientiously. Excellent skills. Asks thoughtful questions. Requires little direction. Keen observer. Goes beyond the required minimum in work. Well prepared. Reports thorough, shows command of principles, and offers logical explanations for discrepancies in data.
6	Very good student. Does all work well but rarely does anything other than assigned. Good skills. Understands work. Shows preparation. Relatively independent, needing little help. Anxious to improve. Good interest. Sincere in his study and work. Reports complete, but discussions of results demonstrate a limited amount of originality of thought.
5	Good student. Completes assignments. Needs occasional help, responds well, corrects mistaken notions and faulty skills, but still uncertain on some points. Skills a little crude. Occasionally unprepared. Moderately interested in work. Reports contain some inaccuracies and omissions, but seem to be the student's own work.
4	Upper average student. Work occasionally incomplete. Needs frequent help, does not always understand readily. Preparation weak. Occasional spurts of interest. Insincere at times. Reports either contain repeated errors and omit items or suggest they are not original work. Data reported for some exercises may well be "dry lab" results.

Figure 3 (cont'd): Teachers' Guide For Assigning Student Grades

ESTIMATED GRADE FOR STUDENT
(KEY)

3	Lower average student. Work frequently incomplete. Low interest level. Seldom prepared. Needs regular help but does not respond well to it. Generally works only when prodded. Poor and incomplete reports and occasionally none at all.
2	Poor student. Work rarely complete. Regularly unprepared. Very low interest. Lab a chore (and bore) to him. Occasionally makes a little effort. Seeks opportunities and means of omitting exercises and assignments.
1	Failing student. Hopeless in the lab. Hardly ever completes work. Very seldom prepared. No interest. Unwilling to work and co-operate. Takes unexcused "cuts".

particular problem focus film. Total scores on these tests were referred to as SRT and PST scores respectively.

It was expected that the COOP test scores would be a good predictor of subsequent performance in carrying out scientific investigations. This expectation followed as a result of the knowledge that the Cooperative General Science Test surveys the concepts in elementary science. It would be reasonable, therefore, to expect that students who achieved well on this test would be better equipped to carry out scientific investigations than those who did not because these test scores depend not only on the knowledge of science but also upon the "ability to apply knowledge in problem situations, and ability to analyze and evaluate scientific ideas and procedures." (25)

The main purpose of including the COOP test scores in the comparisons was to provide a check on teachers' ratings of student performance. In accordance with the argument in the preceding paragraph, the suitability of using teachers' grades as evidence of validity would be questionable if no relationship existed between these grades and scores obtained in the COOP science test.

Hypothesis

> There is no relationship between the scores on the test and scores assigned by teachers to indicate a student's performance in science.

The correlations between teachers' ratings and test scores were computed and the associated probabilities calculated. These results are presented in Table IV.

The hypothesis of no relationship between test scores and teachers' ratings is rejected for five of the six comparisons for the COOP science test, all comparisons for the _Science Reasoning Test_, five of the six comparisons for the _Process of Science Test_, and three out of six comparisons for the _Inquiry Efficiency Test_. Out of a total of 54 comparisons, the null hypothesis is rejected in 47 instances or in approximately 87 per cent of the comparisons.

The test scores which show little or no relationship to ratings given by individual teachers include the following: COOP scores with ratings of teacher E, PST scores with ratings of teacher C, and IET scores with ratings of teachers C, D, and E respectively. With an increase of sample size, it is possible that a relationship would be indicated in all cases except for the comparisons of IET with the ratings of teachers C and D. In these latter two cases the IET is probably measuring something different from what is being evaluated by either of these teachers.

When it seems desirable to average correlation

TABLE IV

CORRELATION OF SCIENCE TEST SCORES AND TEACHERS' RATINGS

TEST	GROUP					
	A	B	C	D	E	X
COOP	0.68**	0.47**	0.35**	0.63**	0.25	0.49**
SRT 1	0.47**	0.39**	0.29**	0.67**	0.58**	0.36**
SRT 2	0.49**	0.33**	0.20	0.34**	0.40**	0.34**
SRT 3	0.49**	0.49**	0.37**	0.43**	0.38*	0.43**
SRT	0.61**	0.54**	0.37**	0.61**	0.63**	0.48**
PST	0.36*	0.27*	0.20	0.60**	0.41**	0.51**
PST 1	0.31**	0.12	0.07	0.53**	0.42**	0.50**
PST 2	0.25*	0.29*	0.23	0.49**	0.26	0.42**
IET	0.43**	0.28*	-0.09	0.01	0.27	0.55**

*$p < 0.05$

**$p < 0.01$

coefficients the Fisher's z statistic is sometimes utilized. Using the transformation as given in Ferguson (32):

$$z_r = \frac{1}{2} \log_e(1 + r) - \frac{1}{2} \log_e(1 - r)$$

where z_r refers to the Fisher's z_r statistic and r the Pearson product-moment correlation coefficient, the correlation coefficients of Table IV were transformed to the corresponding z_r statistic. These were averaged and the corresponding r calculated. The value of t and the corresponding probability was calculated as before. This procedure was used as a device for obtaining further information pertaining to concurrent validity by obtaining an estimate of the correlation between student scores received on the tests and ratings as assigned by teachers. Essentially this meant that an average teacher was postulated. The number of students was also considered in terms of the average number of students per teacher.

As an alternative procedure, the ratings assigned by the teachers were treated as being representative of a larger population of ratings and correlated with the science test scores by using the complete sample as a single group. This resulted in an increase in the number of degrees of freedom and, as expected, each of the comparisons showed a lower associated probability. The results of the comparisons using the two procedures described above are summarized in Table V.

The data in Table V shows that the null hypothesis is rejected for each comparison of the science tests or sub-tests and teachers' ratings. This indicates that an hypothesis of a relationship between the teacher rating and any of the corresponding tests is tenable. To further probe these relationships, the teachers' rating score was selected as the criterion and the scores on the tests as predictors determining the order in which the tests contributed to predictability.

For this purpose a computer program available at the University of Alberta was used. This program calculates a step-wise regression using the method of determinants as outlined by Draper and Smith (30). This method assembles the correlation coefficients into a matrix and selects variables in the regression equation in the order in which the independent variables contribute to the prediction of the dependent variable or response. The first variable Xi is selected on the basis of the ratio $r_{iY}r_{iY}/r_{ii}$ being a maximum. After adjusting the correlation matrix for the entrance of X_i into regression, the next variable is

TABLE V

A SUMMARY OF THE COMPOSITE RELATIONSHIPS BETWEEN SCIENCE TEST SCORES AND TEACHERS' RATINGS

TEST	ESTIMATED COMPOSITE (AVERAGES)			ESTIMATED COMPOSITE (COMPLETE GROUP)		
	r	t	p	r	t	p
COOP	.49	4.4	.001	.47	9.5	.001
SRT 1	.56	5.2	.001	.42	9.2	.001
SRT 2	.36	2.9	.001	.32	6.6	.001
SRT 3	.44	3.7	.001	.45	9.3	.001
SRT	.55	5.1	.001	.52	13.0	.001
PST	.45	3.9	.001	.34	7.0	.001
PST 1	.37	3.0	.001	.27	5.4	.001
PST 2	.35	2.9	.001	.29	5.7	.001
IET	.27	2.1	.03	.21	4.1	.001

selected in the same way using the adjusted matrix and so on.

Tests are made at each stage, by means of F-values, for entry and deletion of a variable from the regression equation. Because of the linear dependency between total scores and sub-scores, predictability was determined first using sub-scores and again using total scores. In the case where the total score was used, the COOP test used as a pretest was included as a predictor. In all cases a probability level of .05 was used for adding or deleting variables.

The results of these procedures are summarized in Tables VI and VII and indicate that the Science Reasoning Test seems to more closely measure the performance measured by teachers' ratings than do either the Process of Science Test or the Inquiry Efficiency Test. When the part that is common to the PST and SRT is accounted for, the IET may be relating to some aspect of the teachers' ratings. The amount of variance accounted for by this particular relationship is too small to make any claims about a definite relationship; i.e. the inclusion of the IET as a predictor increased the amount of variance accounted for by less than one per cent.

TABLE VI

PREDICTION OF TEACHERS' RATING SCORES USING SRT 1, SRT 2, SRT 3, PST 1, PST 2, AND IEF SCORES AS VARIABLES

	Step 1	Step 2	Step 3
Variable Entering	SRT 3	SRT 1	IEF
F value For Variable Entering	93.58	35.16	4.41
Probability	0.01	0.01	0.04
Per Cent Variance Accounted For	20.14	27.07	27.93
Multiple Correlation (R)	0.45	0.52	0.53
F	93.58	68.68	47.68
Probability	0.01	0.01	0.01

Best Regression Equation:

$$TR = 2.000 + .142\ SRT\ 1 + .201\ SRT\ 3 + .019\ IEF$$

TABLE VII

PREDICTION OF TEACHERS' RATING SCORES USING COOP, SRT, PST, AND IEF SCORES AS VARIABLES

	Step 1	Step 2
Variable Entering	SRT	COOP
F Value For Variable Entering	134.29	15.92
Probability	0.01	0.01
Per Cent Variance Accounted For	26.58	29.61
Multiple Correlation (R)	0.52	0.55
F	134.29	77.81
Probability	0.01	0.01

Best Regression Equation:

$$TR = 2.011 + .098\ SRT + .034\ COOP$$

Clearly then, although the hypothesis of a relationship between teachers' ratings and any of the science tests would be tenable, teachers' ratings are more closely associated with the verbal tests than with the tests considered to be less dependent on verbal behavior. This is revealed in Table VII in which approximately 27% of the variance is accounted for in Step 1 using SRT as a predictor and a further 3% added in Step 2 when the COOP is added as the second predictor. In view of the fact that school gradings carried out by teachers are dependent on verbal tests, these results were not disappointing and do not run counter to the expectations that some testing procedures may be measuring abilities not entirely included in verbal tests. Further data related to what was being measured by the science tests is to be found in a later section of this chapter which deals with the overall validity of the testing instruments.

Validity Procedures for HIFAMS

The HIFAMS test was designed as an attitude test but it was believed that the nature of the questions were such that more specific information could be obtained by grouping the questions in such a manner that sub-tests would be formed from questions which seemed to be related to more specific aspects of attitude. Each item was accordingly treated as a single test of a battery of thirty-one tests and the correlation between item scores calculated. Each item was thus considered to represent a separate variable which was related to some aspect of attitude. A principal components factor analysis (50) was carried out in an attempt to resolve this set of variables into a smaller number of categories which could be identified in some meaningful way.

In the mathematical problem carried out by the computer the number of factors printed out was determined by selecting only eigenvalues greater than one from all of the principal components. This limited the number of factors to six. A varimax rotation (50) was carried out and the factors were tentatively identified by noting the items which showed maximum loading on each factor. The HIFAMS test was subsequently divided into six sub-tests formed by grouping together those items which showed maximum loadings on each factor. The solution with varimax rotation is given in Table VIII in which asterisks are used to designate the maximum loading for each test. As presented in Table VIII, items 2, 15, 18, 19, 22, 25, 28, 29, 30, and 31 maximally load on Factor I; items 5, 7, 9, 14, 20, and 26 on Factor

TABLE VIII

FACTOR SOLUTION OF HIFAMS WITH VARIMAX ROTATION

TESTS	COMMUNAL-ITIES	FACTORS					
		I	II	III	IV	V	VI
1.	.661	.231	.083	-.046	.084	.769*	-.013
2.	.467	.399*	.247	.365	.161	.018	-.294
3.	.390	.070	.003	.578*	.181	-.011	.134
4.	.452	.271	.279	.420*	.007	.345	.074
5.	.494	.270	.516*	.224	.217	-.136	.198
6.	.597	-.380	.019	.133	.602*	.134	-.233
7.	.472	.216	.583*	.241	.142	.013	.083
8.	.308	.319	.238	.001	.332*	.198	-.021
9.	.588	.069	.553*	.224	-.108	.424	-.079
10.	.531	.378	.320	.446*	.161	.003	.247
11.	.450	.355	.328	.374*	-.010	.245	-.127
12.	.518	.141	.379	.490*	.123	.098	.303
13.	.414	.176	.204	.577*	.001	.056	-.071
14.	.528	.211	.644*	-.079	.234	.047	-.070
15.	.497	.604*	.199	.233	.024	.180	.075
16.	.460	.344	.306	.455*	-.003	-.006	-.204
17.	.585	.101	.123	.117	-.021	.030	.738*
18.	.369	.393*	.186	.168	.300	.123	-.217
19.	.582	.536*	.471	.021	.123	.059	.233
20.	.496	.084	.542*	.182	.125	.336	.181
21.	.394	.245	.016	.276	.502*	.020	-.067
22.	.514	.430*	.082	.261	.331	.326	.197
23.	.437	.170	-.089	.437*	.243	.366	.128
24.	.563	.243	.196	-.193	.610*	-.137	.194
25.	.494	.598*	.127	.225	.165	.176	-.105
26.	.478	.220	.642*	.128	-.005	.016	.014
27.	.504	.061	.110	.153	.677*	.079	.032
28.	.512	.606*	.171	.122	.228	.189	.116
29.	.579	.696*	.237	.179	.046	.060	.027
30.	.552	.549*	.123	.470	.037	-.050	.105
31	.481	.526*	.316	.131	-.071	.109	.265
Totals	15.339	4.139	3.317	2.825	2.174	1.556	1.327
Per Cent of Common Var.	100	26.98	21.625	18.415	14.175	10.145	8.654
Per Cent of Total Var.	49.48	13.35	10.70	9.11	7.01	5.02	4.28

II; items 3, 4, 10, 11, 12, 13, 16, and 23 on Factor III; items 6, 8, 21, 24, and 27 on Factor IV; item 1 on Factor V, and finally, item 17 maximally loads on Factor VI.

As an aid in identifying the factors, the average value in the correlation matrix used in the analysis was calculated to be $r = 0.23$. This value was used to obtain a rough estimate of the standard error of the factor coefficients using the formula given in Harman (50):

$$a = \frac{1}{2}\,(3/r - 2 - 5r + 4r^2)/N$$

for $N = 373$ and $r = .23$, the standard error is .085. Any coefficient as large as .26 could certainly be regarded as being significantly different from zero.

The items which loaded significantly on one factor and did not load significantly on any other factor were regarded as being keys to the identification problem. The rationale for the identification of each factor is presented below:

Factor I - Items 15, 25, 28, and 29 satisfied the criteria of loading significantly on factor I and not loading significantly on any other factor. These four items have a common theme associated with how well the school staff know and teach the subject matter of the courses available at the school. Factor I was identified as being associated with the students' belief in the proficiency of the school staff.

Factor II - Items 7, 14, and 26 were identified as key items. These items deal with feelings of satisfaction about the school and its facilities. They do not have large loadings on Factor I, already identified as being associated with the proficiency of the school staff, so Factor II was identified as the factor dealing with feelings of satisfaction with school facilities and environment apart from the general proficiency of the school staff.

Factor III - Items 3 and 13 are the key items. These items are associated with a student's feeling of optimism that teachers are interested in him and that he (the student) will achieve future success. Although item 16 does not satisfy both criteria used in choosing key items it loads significantly on Factor III and deals with a feeling of receiving help and assistance. Item 16 tends to confirm the classification of Factor III as dealing with personal satisfaction and gratification.

Factor IV - Items 24 and 27 are key items and deal with a student's opinion of others and how they treat him. Factor IV was therefore identified as being associated with feelings of acceptance by others and was considered to be a sort of congeniality factor.

Factor V - Item 1 serves as the key item. This item deals with the belief in the usefulness of school work. Items 9, 20, 22, and 23 also have significant loadings on this factor. Items 9 and 20 deal with school activities and school spirit respectively, item 22 deals with the usefulness of school work for a satisfying enjoyable life. In the sense that people tend to work at what is purposeful to them, item 23, which deals with the degree to which a student works, could be directly related to the other items. Factor V was therefore identified as the belief in the usefulness of school work.

Factor VI - Item 17 is the key item and deals with feelings associated with the idea of receiving help or guidance from adults. The loading of item 17 on this factor is relatively high (.74) and all other loadings of this item are near zero. Factor VI was therefore identified as the attitude to guidance received from adults.

Sub-tests were formed whose scores were simply the sum of the scores of the items which showed maximum loadings on the respective factors. The sub-tests referred to as HIFAMS(1), HIFAMS(2) etc. were considered to be representative of the corresponding factor whose associations had been identified. This would imply that a high score on a sub-test would be considered as a positive leaning toward the belief, feeling, quality or attitude associated with the factor of the corresponding number. For example, a low score on HIFAMS(1) would be considered to be indicative of some lack of belief in the proficiency of the school staff and so on. Obviously teachers and teaching conditions would vary among school so that negativism and positivism would have meaning mainly within schools.

In order to obtain an estimate of the concurrent validity of HIFAMS and its sub-tests, the test scores were compared to ratings on attitude as assigned by the teachers of the students in the sample. The ratings were essentially rankings within classes and would therefore have little meaning outside the individual class. Any combinations made in order to obtain estimates for the complete sample were done by considering each class as an entity for correlation purposes and then estimating what the correlation might have been had an "average teacher" ranked all students in the complete sample.

Comparing and ranking students in terms of an attribute such as attitude becomes increasingly difficult as the number of comparisons increase. For this reason, cards containing one student name per card were prepared for each class and presented to the teacher in prearranged groups of six cards or less. These prearranged groups were such that each student was ranked within one group and also ranked with respect to at least one individual in every other group. If ratings are done in a consistent manner, it should be possible, in theory, to rank all students from 1 to n where n is the number of students in the class. This should follow directly from the following argument: if A is superior to B and B superior to C, it follows that A must be superior to C. As no ties were allowed for the initial ranking procedure, and each individual was compared directly or by implication with every other individual, a complete ranking would result provided there were no contradictions.

The ranking was accomplished by assigning each student an identification number from 1 to n, where n referred to the number of students in the class. These numbers also corresponded to the row and column numbers of an n by n matrix which was prepared as follows: a "1" was recorded if the student represented by the row number was chosen as being superior to the student represented by the column number and a "0" recorded if the reverse were true. If a tie resulted in the above comparison a "½" was recorded. The rows were summed to produce an attitude score which could have a maximum value of n - 1 and a minimum value of 0. Ties were the result of inconsistencies or the inability of the teacher to distinguish between two individuals. These were resolved with the co-operation of the teachers who carried out further ranking procedures in which the tied ranks were directly compared. In one or two instances ties were allowed to stand because a firm distinction could not be made.

The rankings as described above were used to test the following hypothesis:

Hypothesis:

> There is no relationship between HIFAMS scores and student attitude as rated by their classroom teacher.

In order to obtain an estimate of the correlation between the ranking carried out by the teachers and the test scores, the row sums as described above were treated as scores and an ordinary product-moment correlation calculated. These correlation coefficients were calculated separately for each class and the results summarized in Table IX.

The classes ranged in size from a low of sixteen

TABLE IX

CORRELATION OF HIFAMS TEST RESULTS AND TEACHERS' RATINGS

CLASS	SUB-SCORES						TOTAL	N
	(1)	(2)	(3)	(4)	(5)	(6)		
1.	.27	.34	.07	.15	.46*	-.9	.26	26
2.	.20	.28	.44*	-.07	.32	.05	.34	30
3.	.47*	.12	.55**	.39*	.07	.08	.52**	26
4.	.35	.31	.19	.34	.33	-.13	.36	24
5.	.01	.10	.16	.11	-.07	.01	.09	24
6.	.24	.18	.01	.34	.18	.21	.30	17
7.	.34	.08	.37	-.01	-.02	.14	.33	27
8.	.51*	.34	.52*	.49*	.28	.33	.60*	22
9.	.35	.23	.37	.22	.45*	.42*	.39	21
10.	.04	-.34	.01	.03	.24	-.60*	-.10	17
11.	.33	.29	.41*	.36	.01	-.04	.38*	27
12.	.30	.02	.10	.12	-.14	-.11	.18	21
13.	.39	.38	.20	.28	-.03	-.30	.39	16
14.	.29	.16	.08	.09	.17	-.26	.21	23
15.	.36	.16	.26	.25	.09	-.31	.31	26
16.	.33	.44*	.14	.00	.04	.17	.32	26

*$p < .05$

**$p < .01$

to a high of thirty. These relatively small numbers associated with the calculation of the correlation coefficients meant that the null hypothesis would fail to be rejected for relatively high correlation coefficients; i.e., correlation coefficients would have to have values greater than 0.50 for the samllest class and greater than 0.36 for the largest in order to be considered significantly different from zero ($p < 0.05$). When Table IX is examined it is noted that the correlation coefficients calculated for HIFAMS(6) have an equal number of positive and negative values. Obviously the null hypothesis cannot be rejected for this sub-test. Alternatively, by far the greatest majority of correlation coefficients relating to the other tests have positive values. Out of a total of sixteen correlation coefficients associated with each test, HIFAMS(1) and HIFAMS(3) have no negative correlation coefficients. HIFAMS(4) has two negative coefficients which are both near zero; and HIFAMS(5) has four negative correlation coefficients. Probability considerations suggest that this preponderance of positive values is not due to chance alone.

In order to obtain some crude estimate of what the correlation coefficient would have been had the complete sample been considered as one group, an average correlation coefficient was estimated by first transforming the individual correlation coefficients to the Fisher z_r statistics. The transformed values were averaged and reconverted to correlation coefficients. Although this practice may be somewhat questionable, there is no doubt that the true correlation coefficient for the complete sample would lie within the range of the values for the classes within the sample.

The t value for each of these derived coefficients was determined according to the formula in Ferguson (32):

$$t = r\sqrt{\frac{N - 2}{1 - r^2}}$$

Because of the method used in arriving at the estimated coefficients of correlation, the N used in the above formula refers to the average number of students per teacher (N = 62). The results are summarized in Table X.

The data, as summarized in Table X, indicate that the two sub-tests HIFAMS(5) and HIFAMS(6) are not closely related to teachers' rankings. For each of these two comparisons the null hypothesis is not rejected. HIFAMS(6) shows no relationship whatsoever to teachers' rankings. While HIFAMS(5) may be related to teachers' rankings, the probability level is not sufficiently low to warrant a

TABLE X

A SUMMARY OF THE COMPOSITE RELATIONSHIPS BETWEEN HIFAMS TEST SCORES AND TEACHERS' RATINGS

TEST	ESTIMATED COMPOSITE		
	R	t	P
HIFAMS (1)	.31	2.52	.02
HIFAMS (2)	.21	.67	.10
HIFAMS (3)	.23	1.83	.10
HIFAMS (4)	.21	1.67	.10
HIFAMS (5)	.13	1.01	.20
HIFAMS (6)	-.08	- .50	.20
HIFAMS	.31	2.52	.03

rejection of the null hypothesis. For all other comparisons the null hypothesis may be rejected at the 0.10 level of confidence. From these results it was inferred that HIFAMS and four of its sub-tests are related to what teachers interpret as a good attitude toward school.

Overall Validity of Testing Instruments

In pursuing the problem of what in fact was being measured by the instruments used in this study, it was realized that only a small number of variables had been sampled and used to measure traits and abilities of a particular (and probably non-random) group. The use of factor analysis of the type developed by Kaiser and Caffrey was used in an attempt to make some inferences about the nature of the traits and abilities being tested. Kaiser and Caffrey (58) distinguish between statistical inference and psychometric inference:

> Traditional statistical inference views the N individuals as a (usually random) sample from some larger population and attempts to make inferences about the population from the characteristics of the sample. On the other hand, what might be termed psychometric inference considers the n variables as a (usually non-random) sample from some larger universe of variables, and attempts to infer something about the nature of this universe from a particular selection of n variables. (58)

Theoretically alpha factoring is based on the premise that concern rests with estimating factors existing in a universe of variables when n variables from this universe have been sampled. The solution proceeds by finding $H^{-1}(R-I)H^{-1} + I$, where H^2 are the communalities, R the correlation matrix and I the identity matrix, and determining its r largest eigenvalues and eigenvectors, where r is the number of eigenvalues having values greater than 1. Successive iterations are carried out until H^2 converges within a specified tolerance.

All sub-tests (or complete tests if the sub-tests were not used) were considered to be a selection of variables from the universe of all measuring instruments in the cognitive and affective domain. A factor analysis was performed on these variables in an attempt to make some inference of the nature of the universe of which these variables formed a sample. It was anticipated that the identification of the factors in the resulting factor solution would indicate what was being evaluated, and thus provide further evidence for claims of validity for these measurements.

In the alpha factor solution shown in Table XI, three factors are shown although, for the reasons outlined above, more factors undoubtedly exist. From a table in

TABLE XI

ALPHA FACTOR SOLUTION FOR TWENTY VARIABLES (VARIMAX ROTATION)

VARIABLES	COMMUNALITIES	FACTORS I Verbal	II Affective	III Problem-Solving Ability
I.Q. Verbal	0.59	0.51	0.03	0.57
I.Q. Nonverbal	0.46	0.37	0.08	0.56
TOUS (a) Pretest	0.35	0.59	0.04	0.09
TOUS (b) Pretest	0.39	0.60	0.13	0.16
TOUS (c) Pretest	0.47	0.66	0.14	0.09
TOUS (a) Posttest	0.45	0.63	0.11	0.22
TOUS (b) Posttest	0.54	0.69	0.12	0.23
TOUS (c) Posttest	0.45	0.55	0.22	0.30
COOP Science	0.70	0.71	0.05	0.44

TABLE XI (Cont'd)

VARIABLES	COMMUNAL-ITIES	FACTORS I Verbal	 II Affective	 III Problem-Solving Ability
SRT (1)	0.49	0.63	0.13	0.28
SRT (2)	0.35	0.44	0.20	0.34
SRT (3)	0.31	0.41	0.19	0.34
PST	0.47	0.46	0.17	0.48
IET	0.17	0.17	0.11	0.35
HIFAMS (1)	0.80	0.14	0.89	0.04
HIFAMS (2)	0.57	0.11	0.75	-0.04
HIFAMS (3)	0.63	0.21	0.77	0.03
HIFAMS (4)	0.27	0.08	0.49	0.13
HIFAMS (5)	0.13	0.07	0.35	0.07
HIFAMS (6)	0.15	-0.06	0.24	-0.31
TOTALS	8.73	4.29	2.60	1.84
PER CENT COMMON VAR.	100	49.2	29.8	21.0
PER CENT TOTAL VAR.	43.6	21.5	13.0	9.2

Harman (50) the standard error of a factor coefficient may be obtained as a rough estimate. For an average correlation coefficient r = 0.27, the standard error is 0.072 so any coefficients as large as 0.21 should be significantly different from zero. All of the tests except the Inquiry Efficiency Test and the attitude tests have significant coefficients on the first factor, all of the attitude tests have significant coefficients on the second factor, and all of the tests except the TOUS pretests and the attitude tests have significant coefficients on the third factor. The first factor was identified as a verbal factor, the second as an affective factor and the third as an inquiry or problem-solving ability factor. The third factor was so identified because the nonverbal I.Q. loaded more heavily on it than on the verbal factor and because the science tests (Process of Science Test and the Inquiry Efficiency Test) also favored this factor. The latter two tests had been specifically designed to ensure that the film loops dealt with subject matter which was unfamiliar to the students. From this evidence it was concluded that the third factor was related to performance ability in novel situations.

Clearly then, the universe of variables as inferred from the sample of variables may be characterized by three factors. Two of these were identified as cognitive factors and the third as an affective factor. The verbal factor would seem to be a knowledge factor but the problem-solving ability factor appears to be more directly related to the ability to perform. The affective factor contains only one significant coefficient from the science and the I.Q. tests. This coefficient comes from TOUS(c), which had been classified as dealing with methods and aims of science, and is a relatively low coefficient and thus was considered to be either of little significance or due to some aspects of attitude influencing TOUS(c) posttest scores.

The testing instruments are thus seen to be concerned with three dimensions which may be described as knowledge, performance, and attitude, respectively. It is the performance dimension which is considered to be most closely related to the process aspect of science, therefore statistical procedures were considered necessary in order to minimize the effects of the other two factors in some of the later analyses.

Although only three factors were selected as a consequence of the rule of selecting only factors with associated eigenvalues greater than one, other factors undoubtedly exist. Kaiser and Caffrey (58) point out that, as a consequence of this rule for selecting factors, the number of factors is a function of the number of observable variables. If the number of science tests used in this study were increased, it is probable that some of the uniqueness associated with each test would no longer be unique and the number of eigenvalues greater than one would increase with a corresponding increase in the number of selected factors. It follows that, although the science tests are associated with a verbal factor and a problem-solving ability factor, quite specific competencies peculiar to each test could be present.

C. SUMMARY

1. In the belief that individuals become more proficient in what they practice, the scores on the science tests were considered to reflect proficiencies as follows:

a) The Science Reasoning Test and the Process of Science Test were similar in that each required students to identify processes of science and make judgments about the use of these processes in scientific investigations. Each of these tests was therefore considered to be capable of providing a measure of student knowledge of and skill in the processes of science. This contention received some support from the Pilot Study in which students of a teacher familiar with the Inventory and its purpose achieved higher means in each test than students of a teacher unfamiliar with the Inventory.

b) The score on the Inquiry Efficiency Test (a modification of the TAB Science Test (57)) depends on a student using a minimum number of steps in correctly discovering and verifying hypotheses corresponding to solutions of demonstrated scientific problems. The Inquiry Efficiency Test was therefore considered to be measuring student proficiency in searching for relevant information, processing this information, and discovering or verifying the inference which should be made from a given scientific investigation.

2. The factor solution indicates that the science tests do not have significant loadings on an affective factor, (the one possible exception is TOUS(c) (Understanding About the Methods and Aims of Science)). The science tests are thus seen to be measuring knowledge and performance as

related to the content of each test plus possible specific competencies not revealed by the factor solution.

3. Concurrent validity of the science tests is claimed on the basis that teachers' ratings of student performance in science is significantly related to test scores.

4. Concurrent validity of the HIFAMS attitude test and the sub-tests HIFAMS(1), HIFAMS(2), HIFAMS(3), and HIFAMS(4) is claimed on the basis of the positive relationship between test scores and teachers' rankings of attitude.

5. The Observational Instrument as used by the two observers is reliable in the inter-rater reliability sense.

IV. THE DESIGN OF THE STUDY.

The classical experimental model complete with control groups was considered to be unsuitable for this particular investigation. The experimental curriculum was considered to provide a framework for the actual curriculum, but the actual curriculum under which each group operated had to be determined from data submitted by the teachers during the experiment and from observational procedures. To select control groups in advance would have little meaning other than to provide a comparison between the experimental curriculum and a "traditional" curriculum.

In view of the fact that all of the Junior High School Science Coordinators in the city of Edmonton had some acquaintance with the Inventory and curriculum materials were commonly shared among teachers, curricula which were assuredly not influenced by the Inventory were not available. To achieve any meaningful results with the classical experimental model, it is necessary to (a) randomize groups and assign approximately one-half of the members to the experimental group and the other half to the control group (b) compare the experimental group to the control group prior to treatment (c) give a treatment to the experimental group and no treatment to the control group and (d) compare the experimental group to the control group after treatment. There are many variations of the above four steps but the use of control groups means that one group has a treatment and the other does not. In classroom research dealing with intact classes the assumption of no treatment is almost always untenable. For example, even in the most traditional of science curricula, the process aspects would not be entirely ignored.

In dealing with the comparisons among groups as one of the main concerns of the evaluation model (Fig. 2), it was considered essential to make adjustments in an attempt to offset possible bias inherent in the use of intact classes. Because experimental control is virtually impossible in classroom research, statistical control was attempted by the use of selected concomitant variables in conjunction with a factorial experiment. The factors, in this instance, were the treatment factor and the classification factors of attitude and sex. The concomitant variables were I.Q. scores and COOP science scores. For the TOUS posttest, the TOUS pretest was also used as a concomitant variable. These measures were taken prior to treatment and would be unaffected by treatment. The procedure of using these measurements as concomitant variables was considered to be adequate because their effects were assumed to be linear. The assumptions that treatment effects and regression effects are additive and that residuals are normally and independently distributed with zero means and the same variance were tentatively adopted in the belief that gross violations would be unlikely to occur. Winer (130) states that the additional assumptions underlying the analysis of covariance, over and above that required for ordinary analysis of variance, may be violated to some degree and that the corresponding F tests are robust with respect to these violations. The following chapter deals with the results of these comparisons and with the results of the procedures used to determine the nature of each treatment.

CHAPTER IV

RESULTS

This chapter deals with the testing of hypotheses set out in Chapter I. For organizational purposes it is divided into three main sections. The first section deals with the preliminary data and the associated comparisons in order to ascertain whether the groups differed markedly from one another in science knowledge and tested intelligence as measured by the COOP science test and the I.Q. tests, respectively. The second section deals with the comparisons made among treatment groups after the treatment; while the third section reports the results of the observational procedures and allied analysis and relates these results to the comparisons made among the treatment groups.

I. PRELIMINARY DATA

For this section the hypotheses were not set out formally, but it is to be understood that an alpha level of 0.05 was selected for these preliminary tests and the hypothesis to be tested in each instance would be of the form:

Hypothesis:

> There will be no difference among the means of the treatment groups on the test.

A one way analysis of variance procedure applying the fixed effect model was used for these comparisons. The data were tested for homogeneity of variance using the method outlined by Keeping (60) and the significance of the difference between each pair of means was tested using a Newman-Keuls comparison as outlined in Winer (130). These results are presented in Table XII - XV inclusive.

As seen from Table XII and XIII, the null hypothesis cannot be rejected for differences in treatment groups in the means on the I.Q. tests. The null hypothesis is rejected at the 0.05 level for differences among treatment groups in the means on the COOP science test. A Newman-Keuls comparison reveals that the mean of Group D is significantly higher than the mean of all other groups except Group X ($p < 0.05$). Group D shows some superiority in knowledge and ability in science as measured by the COOP science test. However, this superiority was barely significant ($p = 0.05$). When it is considered that pretesting was done approximately one month after the opening of the school term, there is no real assurance that the difference between Group D and some of the other groups would have been significant had the testing been done prior to the influence due to teacher D.

TABLE XII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE I.Q. (NONVERBAL) TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	108.89	189.09
B	65	106.88	192.68
C	87	105.64	193.21
D	48	110.06	170.05
E	39	108.05	173.43
X	52	107.03	144.75
Total	373	107.59	179.75

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	p
Groups	817	163.40	5	0.91	0.48
Error	66230	180.46	367		

Homogeneity of Variance Test Chi Square = 7.7 p =0.17

TABLE XIII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE I.Q. (VERBAL) TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	106.07	187.16
B	65	105.83	309.53
C	87	103.17	216.85
D	48	108.15	194.48
E	39	107.64	219.09
X	52	106.27	174.88
Total	373	105.81	217.20

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	P
Groups	1013	202.6	5	0.93	0.46
Error	80003	217.99	367		

Homogeneity of Variance Test Chi Square = 6.73 p =0.24

TABLE XIV

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE COOP SCIENCE TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	28.26	92.64
B	65	28.85	86.51
C	87	28.22	93.66
D	48	33.44	85.40
E	39	28.46	95.41
X	52	29.94	125.98
Total	373	29.27	97.14

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	P
Groups	1075	214.94	5	2.24	0.05
Error	35159	95.80	367		

Homogeneity of Variance Test Chi Square = 2.76 p = 0.74

TABLE XV

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS FOR THE COOP SCIENCE TEST

GROUP		D	X	B	E	A	C
	MEANS	33.44	29.94	28.85	28.46	28.26	28.22
C	28.22	5.22*	1.72	0.63	0.24	0.04	0.00
A	28.26	5.18*	1.69	0.59	0.21	0.00	
E	28.46	4.98*	1.48	0.39	0.00		
B	28.84	4.59*	1.10	0.00			
X	29.94	3.50	0.00				
D	33.44						

*Significant, $p < 0.05$

II. COMPARISONS AMONG TREATMENT GROUPS

This section deals with comparisons among groups after the treatment had taken place. The actual differences are first determined using one-way analysis of variance. This is followed by comparisons using controls to remove possible sources of bias inherent in the use of intact groups. The TOUS pretest is included in this section for the sake of clarity and convenience in illustrating comparisons between pretest and posttest.

Tables XVI to XXIII inclusive are devoted to the testing of differences among unadjusted group means. These hypotheses belong to the first group of hypotheses as set out in Chapter I. They are restated here followed by reference to the table in which the corresponding results are reported.

Hypothesis 1.0 There will be no difference among the treatment groups in the means on the HIFAMS attitude test. (Tables XVI-XIX inclusive)

1.1 There will be no difference among the treatment groups in the means on the Process of Science Test. (Table XX)

1.2 There will be no difference among the treatment groups in the means on the Inquiry Efficiency Test. (Table XXII)

1.3 There will be no difference among the treatment groups in the means on the Science Reasoning Test. (Table XXII)

Hypothesis 1.4 There will be no difference among the treatment groups in the pretest means on the TOUS. (Table XXIII)

1.5 There will be no difference among the treatment groups in the posttest means on the TOUS. (Table XXIII)

TABLE XVI

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE HIFAMS(1) SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	33.66	30.67
B	65	34.62	42.02
C	87	36.31	38.78
D	48	41.83	17.59
E	39	40.72	17.21
X	52	38.48	24.49
Total	373	36.91	38.43

ANALYSIS OF VARIANCE

SOURCE	XX	MS	DF	F	P
Groups	3097	619.51	5	20.43	0.001
Error	11238	30.62	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	E	X	C	B	A
A	@	@	@	@	*	
B	@	@	@	*		
C	@	@	@			
X	*	*				
E	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

TABLE XVII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE HIFAMS(2) SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	16.96	12.95
B	65	19.43	13.72
C	87	20.60	14.20
D	48	24.31	7.28
E	39	22.79	6.48
X	72	22.38	8.79
Total	373	20.55	17.08

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	P
Groups	2187	437.47	5	38.37	0.001
Error	4185	11.40	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	E	X	C	B	A
A	@	@	@	@	@	
B	@	@	@	*		
C	@	@	@			
X	@	*				
E	@					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p < 0.05$

TABLE XVIII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE HIFAMS(3) SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	25.59	23.16
B	65	26.74	16.42
C	87	27.82	15.27
D	48	30.52	10.21
E	39	29.69	7.59
X	52	28.96	15.25
Total	373	27.84	18.30

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	P
Groups	1039	207.97	5	13.19	0.001
Error	5785	15.77	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	E	X	C	B	A
A	@	@	@	@	*	
B	@	@	@	*		
C	@	@	*			
X	@	*				
E	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p < 0.05$

TABLE XIX

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE HIFAMS(4) SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	15.91	6.47
B	65	15.80	8.94
C	87	16.62	5.26
D	48	17.83	5.63
E	39	17.23	7.71
X	52	16.46	5.90
Total	373	16.52	6.90

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	p
Groups	167	33.45	5	5.10	0.001
Error	2407	6.56	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	E	X	C	B	A
A	@	@	*	*	*	
B	@	@	*	*		
C	@	*	*			
X	@	*				
E	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

TABLE XX

COMPARISON AMONG THE MEANS OF THE TREATMENT GROUPS ON THE PROCESS OF SCIENCE TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	8.00	9.11
B	65	8.46	9.38
C	87	8.86	8.59
D	48	10.50	10.89
E	39	9.74	8.93
X	52	8.94	13.54
Total	373	8.92	10.32

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	p
Groups	230	45.93	5	4.66	0.001
Error	3619	9.86	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	E	X	C	B	A
A	@	@	*	*	*	
B	@	*	*	*		
C	@	*	*			
X	@	*				
E	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

TABLE XXI

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE INQUIRY EFFICIENCY TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	28.63	57.94
B	65	27.40	72.15
C	87	28.54	60.60
D	48	32.83	29.80
E	39	27.15	51.61
X	52	28.21	48.28
Total	373	28.72	53.90

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	p
Groups	1038	207.57	5	4.00	0.0015
Error	19067	51.95	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	A	C	X	B	E
E	@	*	*	*	*	
B	@	*	*	*		
X	@	*	*			
C	@	*				
A	@					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

TABLE XXII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE SCIENCE REASONING TEST SCORES

GROUP	NUMBER	MEAN	VARIANCE
A	82	14.38	29.99
B	65	14.74	27.10
C	87	14.14	30.10
D	48	18.17	33.46
E	39	15.90	25.52
X	52	16.15	35.11
Total	373	15.28	31.47

ANALYSIS OF VARIANCE

SOURCE	SS	MS	DF	F	P
Groups	654	130.75	5	4.33	0.001
Error	11085	30.21	367		

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

GROUP	D	X	E	B	A	C
C	@	*	*	*	*	
A	@	*	*	*		
B	@	*	*			
E	*	*				
X	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

TABLE XXIII

COMPARISONS AMONG THE MEANS OF THE TREATMENT GROUPS ON THE TOUS SCORES

GROUP	NUMBER	MEAN		STANDARD DEVIATION	
		Pretest	Posttest	Pretest	Posttest
A	82	17.09	17.15	4.72	5.53
B	65	18.94	19.00	4.60	4.05
C	87	18.39	18.38	4.69	5.62
D	48	21.58	22.23	4.02	4.71
E	39	19.79	20.08	3.86	4.72
X	52	19.52	20.35	4.23	4.82
Total	373	18.91	19.18	4.62	5.25

ANALYSIS OF VARIANCE

	PRETEST				POSTTEST			
SOURCE	SS	MS	DF	F	SS	MS	DF	F
Groups	689.13	137.86	5	6.95	976.37	195.27	5	7.69
Error	7284.00	19.85	367	.001	9322.00	25.40	367	0.001

NEWMAN-KEULS COMPARISON BETWEEN ORDERED MEANS

PRETEST

GROUP	D	E	X	B	C	A
A	@	@	@	*	*	
C	@	*	*	*		
B	@	*	*			
X	*	*				
E	*					
D						

POSTTEST

GROUP	D	X	E	B	C	A
A	@	@	@	*	*	
C	@	*	*	*		
B	@	*	*			
E	*	*				
X	*					
D						

@Difference significant, $p < 0.05$

*Difference not significant, $p > 0.05$

4

The results of the foregoing comparisons among treatment groups show that, in terms of the unadjusted means, Group D was superior to: Groups A, C, and B on both the TOUS pretest and the TOUS posttest; Groups A, B, C, and X on the Process of Science Test; Groups E, B, X, C, and A on the Inquiry Efficiency Test; Groups C, A, and B on the Science Reasoning Test. Other significant superiorities ($p \leq 0.05$) were: Group E superior to Group A on the Process of Science Test and both of the Groups E and X superior to Group A on the TOUS (posttest and pretest).

A profile of science achievement depicted in Figure 4 graphically illustrates how treatment groups compared in terms of achievement. The mean for each group was plotted in standard deviation units with reference to the overall mean for each test in order to obtain a comparison of relative achievement. The peak of Group D indicates a definite overall superiority for this group.

These results may have little meaning in isolation because of the possibility of an overwhelming effect of some uncontrolled variables. In the testing of the hypotheses related to the adjusted group means, prior knowledge, I.Q., sex, and attitude were subjected to a measure of statistical control.

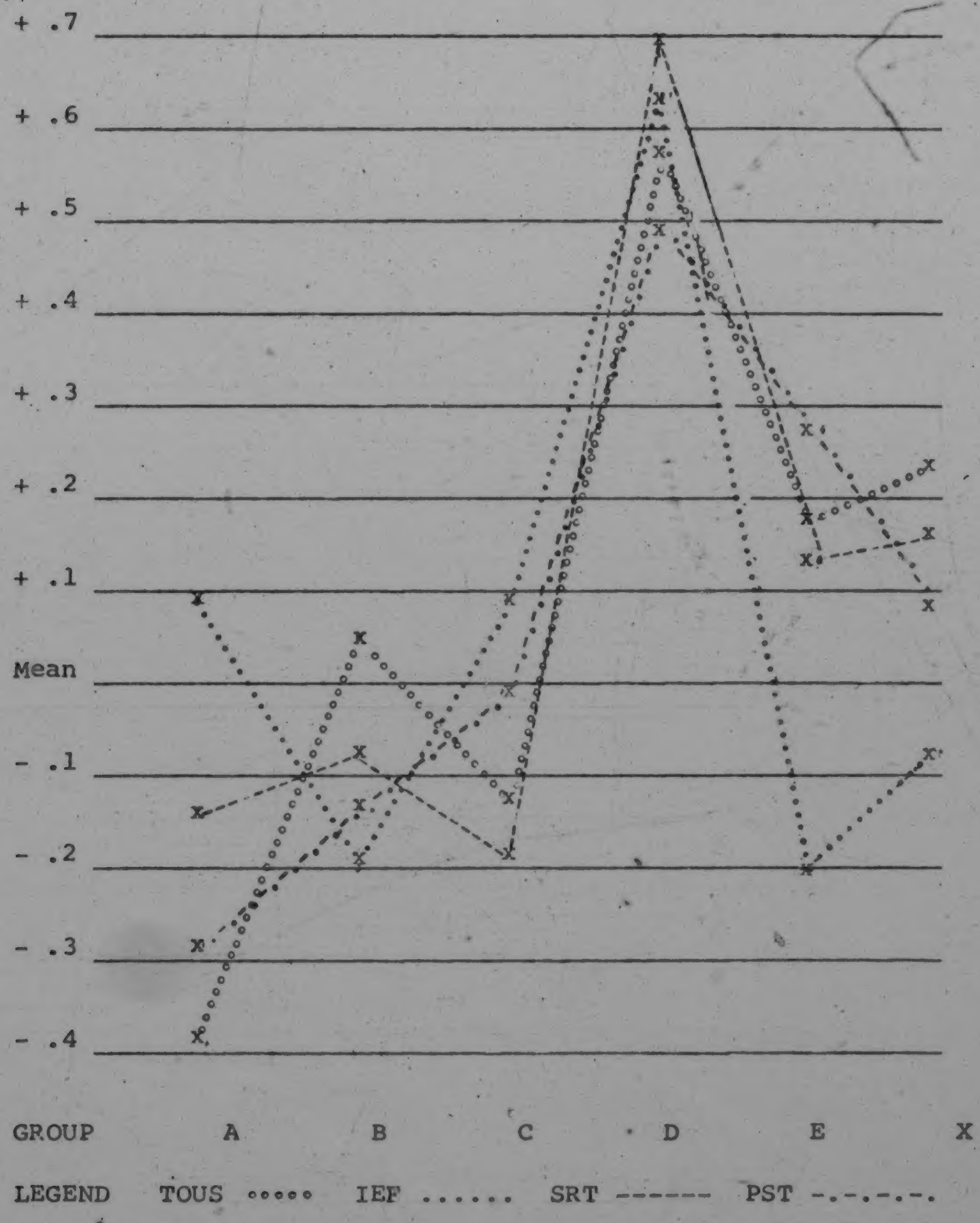


Figure 4. A Profile of Science Achievement for Treatment Groups

Due to cultural expectations for each sex, some subject materials may seem intrinsically more interesting to girls than to boys and vice versa. The attitude of students toward school as related to performance has been investigated in some studies (51, 69). Although the evidence linking performance to positive attitude may not be entirely conclusive, it was considered to be sufficiently strong to warrant exercising some measure of control over this variable.

In the analysis which follows, the factors of sex and of attitude were treated as categorical variables, each with two categories. The attitude categories were determined with reference to the median of each particular group to ensure that groupings with respect to attitude would be within schools. It was recognized that real differences existed among schools, so that a tendency toward a positive attitude would have meaning mainly within individual treatment groups.

A three way analysis of covariance (129) was used for this portion of the analysis. The basic model is of the form:

$$Y_{ijkm}=U+A_i+B_j+C_k+(AB)_{ij}+(AC)_{ik}+(BC)_{jk}+(ABC)_{ijk}$$
$$+d_1Z_{ijkm1}+d_2Z_{ijkm2}+\cdot\ .\ .\ +d_qZ_{ijkmq}+E_{ijkm}$$

$i=1,2,\cdots I$, $j=1,2,\cdots J$, $k=1,2,\cdots K$, $m=1,2,\cdots M_{ijk}$

where

I : no. of levels in factor A (first factor),

J : no. of levels in factor B (second factor),

K : no. of levels in factor C (third factor),

M_{ijk} : no. of observations in (i,j,k) cell,

Y_{ijkm} : mth observation in (i,j,k) cells,

U : overall mean,

A_i : ith level effects of factor A,

B_j : jth level effects of factor B,

C_k : kth level effects of factor C,

$(AB)_{ij}$: interaction effect between ith level of A and jth level of B,

$(AC)_{jk}$, $(BC)_{jk}$, $(ABC)_{ijk}$: interaction effects referred to the corresponding levels,

d_h : regression coefficient on hth concomitant,

q : no. of concomitant variables,

Z_{ijkmh} : hth concomitant variable of mth subject in (i,j,k) cell,

E_{ijkm} : random error or residual of mth subject in (i,j,k) cell.

In all instances, factor A referred to treatments, factor B to attitude, and factor C to sex. These hypotheses are herewith restated followed by reference to the table which reports the corresponding analysis:

Hypothesis 2.0 There will be no difference among the treatment groups in the adjusted means on the Process of Science Test using the I.Q. scores and COOP scores as covariates. (Table XXIV)

Hypothesis 2.1 There will be no difference among the treatment groups in the adjusted means on the Inquiry Efficiency Test using the I.Q. scores and COOP scores as covariates. (Table XXV)

Hypothesis 2.2 There will be no difference among the treatment groups in the adjusted means on the Science Reasoning Test using the I.Q. scores and COOP scores as covariates. (Table XXVI)

Hypothesis 2.3 There will be no difference among the treatment groups in the posttest adjusted means on the TOUS using the I.Q. scores, the COOP scores, and the TOUS posttest scores as covariates. (Table XXVII)

The computer program used in the analyses summarized in Tables XXIV to XXVII inclusive also allowed for determining which linear combinations of parameters were responsible for the significant result if the null hypotheses were rejected at a given alpha level. This researcher was interested only in determining which group differed significantly from what other group and therefore tested for combinations of the form A(i) - A(j), B(i) - B(j) etc. i.e. "Does Group (i) differ significantly from Group (j) in treatment effects, in treatment by attitude effects, etc.?"

TABLE XXIV

COMPARISONS OF TREATMENT GROUPS FOR DIFFERENCES AMONG ADJUSTED MEANS ON THE PROCESS OF SCIENCE TEST

SOURCE	SS	DF	MS	F	P
A Treatments	127.48	5	25.50	3.75	0.01
B Attitude	14.61	1	14.61	0.22	0.64
AB	3.10	5	6.18	0.92	0.47
C Sex	0.06	1	0.06	0.01	0.93
BC	4.06	1	4.06	0.60	0.44
AC	23.27	5	4.65	0.69	0.64
ABC	26.40	5	5.28	0.78	0.57
Cov. 1 I.Q. Verbal	19.74	1	19.74	2.91	0.09
Cov. 2 I.Q. Nonverbal	31.86	1	31.86	0.47	0.03
Cov. 3 COOP Science	322.20	1	322.20	47.45	0.001
Error	2349.57	346	6.79		

TABLE XXIV

COMPARISONS OF TREATMENT GROUPS FOR DIFFERENCES AMONG ADJUSTED MEANS ON THE PROCESS OF SCIENCE TEST

SOURCE	SS	DF	MS	F	p
A Treatments	127.48	5	25.50	3.75	0.01
B Attitude	14.61	1	14.61	0.22	0.64
AB	3.10	5	6.18	0.92	0.47
C Sex	0.06	1	0.06	0.01	0.93
BC	4.06	1	4.06	0.60	0.44
AC	23.27	5	4.65	0.69	0.64
ABC	26.40	5	5.28	0.78	0.57
Cov. 1 I.Q. Verbal	19.74	1	19.74	2.91	0.09
Cov. 2 I.Q. Nonverbal	31.86	1	31.86	0.47	0.03
Cov. 3 COOP Science	322.20	1	322.20	47.45	0.001
Error	2349.57	346	6.79		

TABLE XXV

COMPARISONS OF TREATMENT GROUPS FOR DIFFERENCES AMONG ADJUSTED TREATMENT MEANS ON THE INQUIRY EFFICIENCY TEST

SOURCE	SS	DF	MS	F	p
A Treatments	787.47	5	157.49	3.46	0.01
B Attitude	110.41	1	110.42	2.42	0.12
AB	597.92	5	119.59	2.62	0.02
C Sex	18.92	1	18.92	0.42	0.52
BC	19.15	1	19.15	0.42	0.52
AC	28.15	5	5.63	1.24	0.29
ABC	169.34	5	33.87	0.74	0.59
Cov. 1 I.Q. Verbal	1.93	1	1.93	0.04	0.84
Cov. 2 I.Q. Nonverbal	478.13	1	478.13	10.50	0.001
Cov. 3 COOP Science	130.75	1	130.75	2.87	0.09
Error	15760.70	346	45.55		

TABLE XXVI

COMPARISONS AMONG TREATMENT GROUPS FOR DIFFERENCES AMONG ADJUSTED TREATMENT MEANS ON THE SCIENCE REASONING TEST

SOURCE	SS	DF	MS	F	P
A Treatments	133.16	5	26.63	1.58	0.17
B Attitude	26.08	1	26.08	1.55	0.22
AB	9.68	5	19.31	11.45	0.34
C Sex	16.36	1	16.36	0.97	0.33
BC	14.23	1	14.23	0.84	0.36
AC	61.09	5	12.22	0.74	0.61
ABC	74.09	5	14.82	0.88	0.50
Cov. 1 I.Q. Verbal	84.11	1	84.11	4.98	0.03
Cov. 2 I.Q. Nonverbal	95.63	1	95.63	5.67	0.02
Cov. 3 COOP Science	1429.77	1	1429.77	84.77	0.001
Error	5835.88	346	16.87		

TABLE XXVII

COMPARISONS AMONG TREATMENT GROUPS FOR DIFFERENCES AMONG ADJUSTED TREATMENT MEANS ON THE TOUS (POSTTEST)

SOURCE	SS	DF	MS	F	p
A Treatments	200.81	5	40.16	3.47	0.01
B Attitude	17.52	1	17.52	1.51	0.22
AB	113.92	5	22.78	1.97	0.08
C Sex	30.05	1	30.05	2.60	0.11
BC	19.97	1	19.97	1.73	0.19
AC	64.79	5	12.96	1.12	0.35
ABC	65.04	5	13.01	1.12	0.35
Cov. 1 I.Q. Verbal	43.09	1	43.09	3.72	0.05
Cov. 2 I.Q. Nonverbal	4.64	1	4.64	0.40	0.53
Cov. 3 COOP Science	604.39	1	604.39	52.23	0.001
Cov. 4 TOUS pretest	525.70	1	525.70	45.43	0.001

The data in Tables XXIV to XXVII inclusive indicate that there were no statistically significant effects due to sex or attitude ($p < 0.01$). It was therefore concluded that treatments were relatively unaffected by the categorical variables of attitude or sex. It was further concluded that hypothesis 2.2 fails to be rejected, but hypotheses 2.0, 2.1, and 2.3 are not accepted. These results are discussed below.

After allowing for initial differences among the treatment groups by means of covariance adjustment from measurements of intelligence, as measured by I.Q. tests, and Science knowledge, as measured by the COOP science test; there was no significant difference among the treatment groups in the adjusted means on the Science Reasoning Test. The analysis of variance done on the COOP, Table XIV, indicated that the groups were significantly different on this covariate ($p < 0.05$) and in view of the fact that pretesting was done approximately one month after the opening of the school term, there may be some probability that the use of the COOP science test as a covariate may have removed some of the effects of the teaching procedures or treatments. It was concluded, however, that there was no significant difference among groups in their knowledge of and skill in the processes of science as measured by the *Science Reasoning Test*.

Significant differences were found among groups in the adjusted means on the *Process of Science Test*, the *Inquiry Efficiency Test*, and the TOUS posttest ($p < 0.01$) in each case. The *Process of Science Test* was considered to be less dependent on verbal skills than the *Science Reasoning Test* and the removal of some of the verbal effects may have had some bearing on the fact that a significant difference was found in the adjusted means of the PST but not on the adjusted means of the SRT. It was concluded that after allowing for initial differences among treatment groups in tested I.Q. and knowledge of science, there were significant differences among the six groups in their (1) knowledge of and skill in the processes of science (as measured by the PST), (2) efficiency in inquiry procedures; i.e., the ability to select questions pertinent to the solution of a scientific problem or investigation (as measured by the IET), and (3) knowledge about scientists and the scientific enterprise (as measured by the TOUS).

A comparison of the adjusted treatment means on the science tests is presented in Table XXVIII.

TABLE XXVIII

A COMPARISON OF ADJUSTED TREATMENT MEANS ON THE SCIENCE TESTS

		TESTS			
GROUP	N	TOUS	SRT	PST	IET
A	82	18.08	14.57	8.09	28.55
B	65	18.93	14.90	8.54	27.52
C	87	19.02	14.68	9.12	28.87
D	48	20.75	16.94	9.95	32.26
E	39	19.84	15.99	9.22	27.03
X	52	19.99	15.97	8.82	28.22

It is clear that Group D remained in a superior position relative to the other groups although the magnitude of this superiority has been somewhat reduced as a result of the covariance adjustments.

A feature of the computer program which allowed for testing for the significance of the difference between any pair of means was utilized to determine which of the mean differences contributed to the significant result for the PST and for the IET. All possible combinations were tested. The results of the tests which were found to be significant are printed in Table XXIX. The test is essentially a Hotelling T^2 test (77) and outputs estimated difference between means, lower and upper bounds of difference, variance, and probability level.

Group D achievement, on the basis of the adjusted means, is thus significantly superior to that of Group A on the Process of Science Test and significantly superior to that of Groups B and E on the Inquiry Efficiency Test.

Although significant differences existed among groups in the adjusted means of the TOUS, the mean of Group D was not significantly superior to the mean of any other

TABLE XXIX

SIGNIFICANT DIFFERENCES IN ADJUSTED GROUP MEANS (PROCESS OF SCIENCE TEST AND INQUIRY EFFICIENCY TEST)

Instrument	Comparison of Group Mean Difference	Lower Bound	Upper Bound	Estimated Difference Between Means	Variance	p
PST	D - A	0.03	0.35	1.75	0.26	0.04
IET	D - B	0.51	9.93	5.22	1.98	0.02
	D - E	0.32	10.60	5.47	2.36	0.03

single group. It was realized that changes had occurred and an attempt was made to find out the nature of the changes as indicated by significant shifts from a wrong answer on the pretest to a right answer on the posttest (or vice versa). It was anticipated that trends of success would be indicated by significant shifts from wrong answers to right answers to portray a growth pattern. Groups could therefore be compared in terms of fairly specific questions such as, "Which group has shown the greatest trend of success in the portion of the TOUS classified as Understanding About the Methods and Aims of Science?"

In view of the fact that Groups E and X were in the same school and were combined on occasions plus the fact that their adjusted means on the TOUS were practically equal, these groups were combined as one group referred to as Group E + X for subsequent analysis on the TOUS. In order to provide a sort of standard for what might be termed "expected trends," the total sample was included as a group for comparative purposes.

In ascertaining the number of items for which a significant shift occurred between the TOUS pretest and the TOUS posttest, the following data were obtained for each item of the TOUS:

1. The number of students in each treatment group obtaining a correct answer for the pretest (x)
2. The number of students in each treatment group obtaining a correct answer for the posttest (y)
3. The number of students in each treatment group obtaining a correct answer for both the pretest and the posttest (q)

From these data a normal deviate was calculated from the formula:

$$z = \frac{y - x}{\sqrt{x + y - 2q}}$$

This formula follows almost directly from the formula in Ferguson (32):

$$z = \frac{D - A}{\sqrt{A + D}}$$

Where D and A refer to the cell frequencies as illustrated:

		2nd	
		FAIL	PASS
1st	PASS	A	B
	FAIL	C	D

The results of these procedures are presented in Tables XXX and XXXI.

TABLE XXX

NORMAL DEVIATES FOR TOUS TEST ITEMS ESTIMATING CHANGE BETWEEN PRETEST AND POSTTEST

ITEM			GROUP		
	A	B	C	D	E+X
1.	1.00	1.89	2.31*	-0.58	0.30
2.	-1.06	-0.57	- .33	-0.58	-0.38
3.	0.24	-1.00	-0.2	2.65*	0.91
4.	-0.17	-0.69	1.35	-0.78	1.35
5.	0.34	-1.51	-0.507	2.06*	-0.71
6.	-0.63	0.65	-1.40	2.45*	0.93
7.	0.78	0.47	1.80	2.14*	0.22
8.	-0.60	0.00	0.39	-0.30	0.00
9.	1.83	1.09	0.73	1.70	2.20*
10.	0.71	0.58	-1.26	2.43*	3.28*
11.	-1.16	-1.51	-1.51	1.07	0.23
12.	-0.87	2.24*	0.37	1.96*	1.03
13.	-0.34	0.43	1.06	-0.69	-0.37
14.	0.54	0.78	0.00	0.00	-0.87
15.	0.94	2.29*	0.17	1.00	1.42
16.	-0.74	-1.15	0.54	2.00*	1.67
17.	-3.55	0.54	-1.15	-1.34	1.62

*Significant, $p < 0.05$

@Significant change from right answer to wrong answer, $p < 0.05$.

5

TABLE XXX (continued)

ITEM	GROUP A	B	C	D	E+X
18.	-1.09	-0.19	-1.37	0.471	-1.03
19.	0.200	0.54	3.09*	-2.36@	-1.57
20.	1.40	1.51	0.49		0.97
21.	2.84*	-0.30	0.47	2.45*	1.15
22.	-1.20	1.10	-1.62	0.22	-2.29@
23.	0.54	0.54	0.16	2.36*	-0.51
24.	0.00	0.85	0.00	-2.11@	1.83
25.	1.44	-1.46	0.38	0.45	-0.90
26.	1.25	0.28	-0.58	1.39	0.90
27.	0.34	-1.57	-1.98@	-3.00@	-3.16@
28.	1.06	-1.79	-1.18	1.41	1.88
29.	-0.94	-1.07	-1.35	1.13	3.00*
30.	-3.29@	0.19	0.17	-1.21	-0.60
31.	1.35	-0.23	0.35	1.73	1.96*
32.	1.67	0.28	3.14*	1.00	-1.15
33.	1.42	0.41	-0.16	1.97*	0.76
34.	-0.69	-0.78	0.17	-1.41	-0.34
35.	0.96	-1.18	-1.57	0.00	1.28
36.	-0.66	-1.61	0.50	-1.09	-1.04

*Significant, $p < 0.05$

@Significant change from right answer to wrong answer, $p < 0.05$

TABLE XXXI

NORMAL DEVIATES FOR TOUS TEST ITEMS ESTIMATING CHANGE BETWEEN PRETEST AND POSTTEST FOR THE COMPLETE SAMPLE

ITEM	GROUP COMPLETE SAMPLE	ITEM	GROUP COMPLETE SAMPLE
1.	2.50*	19.	0.18
2.	-1.22	20.	1.75
3.	0.93	21.	2.76*
4.	0.64	22.	-1.86
5.	-0.33	23.	1.14
6.	-0.64	24.	0.60
7.	2.02*	25.	0.09
8.	-0.20	26.	1.40
9.	3.32*	27.	-4.00@
10.	1.81	28.	0.53
11.	-1.47	29.	0.
12.	1.76	30.	-1.81
13.	0.01	31.	2.18*
14.	0.02	32.	3.68*
15.	2.48*	33.	1.85
16.	0.99	34.	-1.24
17.	-1.89	35.	-0.28
18.	-1.57	36.	-2.20@

*Significant, $p < 0.05$

@Significant change from right answer to wrong answer, $p < 0.05$

In interpreting the data of Tables XXX and XXXI, two considerations must not be overlooked. For items which a large number of students answered correctly on the pretest, it is difficult to find any change that is statistically significant in terms of improvement; also a statistically significant change may occur for items which are extremely difficult. Klopfer and McCann (61), in a similar situation, adopted the criteria that no item was to be considered an indication of weakness unless it was not statistically significant and fewer than eighty per cent of the students responded to it correctly on the pretest in each year in which the test was given. At the other end of the scale, an item was not used as an indicator of success if fewer than twenty per cent of the students answered it correctly on the posttest. From these considerations, item 21 and item 36 were not considered to be good indicators of a trend because fewer than twenty per cent of the students answered these items correctly on the posttest.

Forty-six per cent of the students selected response "A" and only nineteen per cent selected the correct response "C" for item 21. This could be an attitude effect due to some tendency on the part of younger students to consider scientists as being directly associated with the application of all scientific knowledge. This item as illustrated below:

21. Different groups of people help mankind in different ways. What is the special way in which scientists help mankind?

    A. Scientists make better things for better living.
    B. Scientists show us how to be more healthy.
    C. Scientists give us knowledge about nature.
    D. Scientists offer skilled service and advice.

This item may serve to discriminate between a belief in the "altruistic scientist" and a more thorough understanding of the methods and aims of science, but for statistical reasons was not considered a good indicator of trend for this particular sample.

For the other example:

36. "Most scientists are smart. They learn more easily than most people and can do harder things with their minds."
    Is this statement correct?

    A. Yes, but scientists are generally no smarter than doctors or lawyers.
    B. Yes, but only because scientists are born with scientific skills.
    C. Yes, but only because scientists have received special training.
    D. No. Scientists are about as smart as most people but no smarter.

Response "C" was favored by fifty per cent of the group. The failure of many students to realize that the word scientist applies to a multitude of people and not just to the prominent men of science may have led to the failure to select the correct response "A".

All groups, except group A, show a change from right to wrong answer on item 27. Each of groups C, D, and E show a significant change.

27. Scientists study plants mainly to:

A. help farmers to produce more food.
B. discover how to make new medicines.
C. understand how they live and grow.
D. find out where they will grow best.

An examination of this item reveals no ambiguity. It is possible, however, that these young students, engaged in a study of biological science, interpreted the correct response (C) to mean understanding in a broad sense. The level of sophistication of this age group could conceivably have led some students to reject the correct response because they had become aware that scientists understand a great deal about how plants live and grow. Essentially, the shift from right answer to wrong answer may have been part of a learning pattern in which the student rejects the accepted answer because he has a greater understanding than he had initially but an insufficient understanding to realize the complexities involved in the life and growth of plants. This argument, coupled with the fact that the Groups (D and E+X) whose means exceeded that of the total sample showed a significant change from right to wrong answer, suggests that this particular item should be omitted.

In Tables XXX and XXXI, the negative sign indicates a change from right answer to a wrong answer. All groups show a greater frequency of change from wrong to right than vice versa but this frequency is not an overwhelming one. In terms of significant change per item only group D seems to have achieved any notable improvement. When the total sample is considered, some positive changes become statistically significant and indicate some overall measure of success. The following summary serves to illustrate the growth (or lack of growth) patterns.

GROUP	ITEMS FOR WHICH A SIGNIFICANT SHIFT OCCURRED	
	WRONG TO RIGHT	RIGHT TO WRONG
A		17,30
B	12,15	
C	1,19,32	
D	3,5,6,7,10,12,16,23,33	19,24
E+X	9,10,29,31	22
TOTAL	1,7,9,15,31,32	

In an attempt to obtain further evidence about the differences among groups, the foregoing data were used in conjunction with the considerations below.

If no net change existed between pretest and posttest results, there would be approximately the same number of positive shifts as there were negative shifts. Assuming that the changes were randomly distributed among the items, statistically significant changes should occur in 5% of the items (2 items or less). That is, there should be 1 item in which there was a significant positive change and 34 items without significant positive changes. If under the assumption of random distribution of changes it can be demonstrated that a greater number of positive shifts occurred than could be attributed to chance, a contradiction exists. It may then be concluded that the assumption of random distribution is untenable and/or there are a greater number of positive shifts than could be attributed to chance alone.

The actual results of each group (in terms of items with statistically significant positive shifts) were compared to the expected results by means of a chi square test. Only in the case of group D was any significant difference found. For group D, $\chi^2 = 7.5$ and χ^2 (with Yates's correction) = 5.7. χ^2 is clearly significant in each case.

Clearly then, group D has shown the greatest positive change and group A the least. For group D, three positive and one negative change belong to the portion of TOUS classified as _Understanding About the Scientific Enterprise_, two positive changes belong to _Understanding About the Personal Characteristics of Scientists_ and four positive changes and one negative change belong to the portion classified as _Understanding About the Methods and Aims of Science_. These sections have been heretofore classified as TOUS(a), TOUS(b), and TOUS(c), respectively and the remaining shifts under this classification are as follows:

TOUS(a) - E+X (3), Total (1)
TOUS(b) - B(1), E+X (1), Total (1)
TOUS(c) - A (-2), B (1), C (3), Total (4)

in which the number in brackets immediately following the group classification refers to the total number of items and the negative prefix indicates that the change is from a right answer to a wrong answer.

Although all groups operated under the same overall curriculum, subject to the same behavioral objectives as implied by the _Inventory_, there were significant differences in achievement among groups. After allowing for initial differences in I.Q. and in general science achievement. Group D had attained greater knowledge of and skill in the processes of science than Group A and greater efficiency in inquiry procedures (ability to select questions pertinent to the solution of a scientific investigation) than either Group B or Group E. Although Group D was not shown to be significantly superior to any other single group in the knowledge about science and scientists as measured by the adjusted means on the TOUS (posttest), the group showed a greater positive shift from wrong answers to right answers on the TOUS than any other group. It is suggested therefore that the treatment received by Group D was positively related to achievement in the process dimension. The remainder of this chapter is devoted to the determination of the kinds of treatment received by each group and with relating of treatments to achievement.

III. RESULTS OF OBSERVATIONAL PROCEDURES AND ALLIED ANALYSES*

The information obtained from the reports submitted by the participating teachers was summarized in terms of the time devoted to each instructional mode and to each process (Tables XXXII and XXXIII). Some difficulty was encountered in interpreting the meaning which could be attached to the summary because of individual variations as to the meaning attached to a given instructional mode. For example, simulation as reported by teacher B, referred to the student acting as a scientist in dealing with particular investigations. The other participants generally placed this activity within the laboratory category.

---

*Data in Tables XXXII to XXXV inclusive were obtained jointly with Powley (94). They are repeated here in the interests of clarity.

The time devoted to a given instructional mode seems quite definitely more related to teaching style than to the availability of resources or facilities. Teacher A, for example, had laboratory facilities which were inferior to those available to teacher B yet he devoted far more time to this particular mode, (40.1% as opposed to 1.8%). Even when the time devoted to simulation is added to the time devoted to laboratory for teacher B, the time ratio of laboratory time to total time is almost four to one in favor of teacher A, (40.1% to 10.4%).

It had also been anticipated that some information

TABLE XXXII

A COMPARISON OF TEACHERS IN TERMS OF PERCENTAGE OF TIME DEVOTED TO DIFFERENT INSTRUCTIONAL MODES

INSTRUCTIONAL MODES	TEACHERS				
	A A	B	C	D	E
Laboratory	40.1*	1.8	32.3	33.8	36.3
Demonstration	0.0	3.7	5.0	14.5*	0.0
A-V Media	2.7	23.2*	14.8	0.0	21.6
Field Study	9.5*	3.5	0.0	0.0	4.7
History of Science	0.0	0.0	0.0	0.0	2.1
Invitations to Inquiry	0.0	9.2*	0.0	0.0	0.0
Simulations	2.7	8.6*	1.4	0.6	0.0
Project	0.0	0.9	0.0	0.0	6.9*
Library Research	7.5	12.1*	0.0	0.0	0.0
Lecture	6.9	6.8	1.2	21.5*	20.4
Discussion	23.9	25.0*	18.7	12.9	17.6
Other@	6.8	5.3	29.4*	14.9	0.0

*Largest value for particular mode

@Other: includes drill, work sheets, chart making, etc. not specifically listed under instructional modes.

TABLE XXXIII

A COMPARISON OF TEACHERS IN TERMS OF PERCENTAGE OF TOTAL PROCESSES SPENT IN EACH GROSS PROCESS AND IN EACH MAJOR PROCESS

PROCESS	TEACHERS				
	A	B	C	D	E
I. INITIATION	27.7	52.1	38.2	47.8	41.2
1. Identifying Problems	2.7	11.7	1.9	17.4	6.7
2. Background Information	7.4	24.2	9.1	9.9	20.0
3. Predictions	6.1	10.8	13.8	0.4	1.2
4. Hypotheses	4.7	4.6	10.5	5.1	5.4
5. Design of Experiment	6.8	0.8	2.8	15.0	7.9
II. COLLECTION OF DATA	27.0	10.4	29.0	24.5	25.8
6. Procedure	8.8	0.0	13.1	7.5	12.5
7. Observations	18.2	10.4	15.9	17.0	13.3
III. PROCESSING OF DATA					
8. Organizing Data	16.9	13.3	3.1	4.7	15.0
9. Representing Data Graphically	4.0	5.4	3.1	2.4	3.3
10. Mathematical Treatment	1.4	1.2	1.1	0.8	0.0

TABLE XXXIII (continued)

PROCESS	TEACHERS				
	A	B	C	C	E
IV. CONCEPTUALIZATION OF DATA	10.8	16.3	20.5	13.1	4.6
11. Interpreting of Data	6.8	15.1	16.8	11.9	2.5
12. Operational Definitions	8.0	0.0	3.1	1.2	2.1
13. Mathematical Relationships	0.0	1.2	0.0	0.0	0.0
14. Mental Models	0.0	0.0	0.6	0.0	0.0
V. OPEN-ENDEDNESS	12.1	8.4	6.0	9.0	10.0
15. Further Evidence	4.0	6.7	0.0	2.4	0.4
16. New Problems	0.0	0.0	6.0	3.6	4.6
17. Application	8.1	1.7	0.0	2.8	5.0

would have been forthcoming in terms of the sequence in which instructional modes were used to reveal possible individual patterns. The nature of the actual reports received and the omission of datings, in some instances, made this impossible. Nevertheless certain patterns clearly emerge in terms of teacher preference for certain instructional modes. Similarly, patterns appear to exist for teacher emphasis upon particular processes as set out in the Inventory.

The observational instrument, referred to in Chapter III, was used in order to obtain a description of the short term teaching procedures that are involved within one classroom period and to make judgments about the patterns emerging from the long term procedures. This researcher was particularly interested in the actual meanings attached to the various processes of science as interpreted by each individual teacher. Were these interpretations identical with or similar to that held by the researcher and his colleague?*

---

*See Chapter III p.119 for method used to establish agreement between observers and thus form the "criterion" for the Observational Instrument.

The formal observations occupied three regular classroom periods for each active participant or a total of approximately 120 minutes of recording time per teacher. The percentages of time, devoted to each process and to each teaching act were calculated and the results summarized in Table XXXIV.

The criterion, as established from the agreement between the two observors on the seventeen gross processes, was used to test Hypothesis 3.0 which is restated below:

Hypothesis 3.0 There will be no difference between the teacher's perception, of the science process he is emphasizing, and the science process he is actually emphasizing.

TABLE XXXIV (PART I)

PERCENTAGE OF TIME DEVOTED TO EACH PROCESS DURING FORMAL OBSERVATION

PROCESS	TEACHERS				
	A	B	C	D	E
1. Identifying and Formulating a Problem	8.2			5.6	1.9
2. Background Information	8.2	62.9	7.1	4.7	9.8
3. Predicting					
4. Hypothesizing	18.9		4.0		
5. Design	13.5				3.9
6. Procedure	21.6			40.1	17.6
7. Observations	21.6	29.0	49.5	30.8	42.2
8. Organizing of Data		6.4	2.0		3.9
9. Representing the Data Graphically					3.9
10. Treating Data Mathematically		1.6	37.4		16.7
11. Interpreting the Data	8.2			13.1	
12. Form Operational Definitions					
13. Mathematical Relationships					
14. Theory Building					
15. Further Evidence					
16. New Problems					
17. Applying the Knowledge					

TABLE XXXIV (PART II)

PERCENTAGE OF TIME SPENT IN DIFFERENT TEACHING ACTS DURING THE FORMAL OBSERVATION

	TEACHING ACTS	TEACHER				
		A	B	C	D	E
TEACHING TECHNIQUES	**Teacher Talks**					
	Gives Directions	2.7	3.2	3.4	3.8	18.6
	Introduces	3.9				
	Lectures	2.7				
	Summarizes	1.3		2.3		2.1
	Explains	20.8		1.2		6.9
	Teacher-Student Talk					
	Recitation				5.7	
	Request and Ans. Quest.	29.9		29.9	5.7	9.3
	Discussion	9.3		2.3	9.4	1.4
	Teacher Does					
	Uses A-V			36.8		4.2
	Demonstration	3.9				8.4
	Helps Indiv. Student[@]	4.3	44.2	12.6	35.8	16.8
	Students Do					
	One Student-Class					0.7
	Students Work Individ.		40.0	11.5		6.3
	Laboratory Work	20.8	12.6		39.6	23.8

[@] Helping of individual students took place during times when students were engaged in laboratory work or when others were working independently.

TABLE XXXIV (PART II) (continued)

	TEACHING ACTS	TEACHERS				
		A	B	C	D	E
	Purpose					
	Review		1.6	13.3		
	Evaluation					
	Teacher's Questions@					
	Recall Facts	10.0		35.6	19.1	37.5
	See Relationships	50.0		35.6	38.1	25.0
	Make Observations	30.0		20.0	28.6	37.5
	Hypothesize	10.0		8.9	14.3	
	Test Hypotheses					
CHARACTERISTICS OF TEACHING	Method					
	Concrete	70.0	25.0	56.3	63.4	64.7
	Abstract	29.9	75.0	43.8	26.6	35.3
	Subject Matter					
	Practical	22.2	2.7	36.4	0.0	24.2
	Theoretical	77.8	97.8	63.6	100.0	75.8
	Pupil Activity					
	Directed	100.0	97.9	100.0	100.0	100.0
	Non-Directed		2.1			

@No questions audible for teacher B

Each teacher was told prior to each lesson that he would be asked to list the processes with which he had dealt during the course of the lesson. It was assumed that this practice of giving prior notice would minimize possible errors due to memory effects.

The null hypothesis was rejected for any case in which the teacher's selection failed to include at least 60% of the processes selected by the criterion or if the teacher's selection could be attributed to chance alone ($p < .10$). Each selection that was 60% or more in agreement with the criterion was tested using the following considerations from elementary probability as described below.

Given that a teacher selected t processes and r of these processes were in agreement with the criterion. Let T be the number of possible samples of t processes which include r or more of the processes selected by the criterion and let C be the number of possible samples of t processes included in the 17 processes. The probability (p) would be:

$$p = \frac{T}{C}$$

For the total observational period the probability would be:

$$p = \frac{T1 + T2 + T3}{C1 + C2 + C3}$$

The numbers 1, 2, and 3 refer to the first, second, and third observational period respectively.

The probability calculations for the teachers' selections gave the following results:

Teacher	Probability
A	.09
B	.12 (reject)
C	.13 (reject)
D	.02
E	.01

The null hypothesis is thus rejected for teacher B and for teacher C. It cannot be claimed, therefore, that agreement with the criterion was achieved for either of these teachers. Consequently, the summaries prepared by them may not be entirely credible and should be subject to rather cautious interpretations.

According to Table XXXV, each of these teachers perceived that he was dealing with Identifying Problems; yet his actions in this regard were not identified as

TABLE XXXV

SUMMARY OF TEACHER PERCEPTION OF PROCESSES TAUGHT COMPARED TO CRITERION

PROCESSES	Teacher A			Criterion			Teacher B			Criterion			Teacher C			Criterion			Teacher D			Criterion			Teacher E			Criterion		
	LESSONS																													
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Identifying Prob.				x			x	x						x	x				x	x	x	@	@	@						x
Background Info.	x	x		@	@		x	x	x	@	@	@	x	x	x	@				x	x	x		@	x	x	x	@	@	@
Predictions													x	x	x															
Hypotheses	x	x		@	@	x							x	x	x			@												
Design				x	x														x	x	x	@	@				x			@
Procedure	x	x		@	@								x						x	x	x	@	@		x	x	x	@		@
Observations	x	x	x			@			x		x	@	x	x	x	@	@	@		x	x	x	@	@	x	x	x	@	@	@
Organizing Data	x		x									x	x			@									x			@		
Representing Data Graphically																													x	
Mathematical Treatment																									x	x		@	@	
Interpreting Data			x			@							x	x	x	@	@	@						x						
Operational Def.														x	x															
Math. Relationship																														
Theory Building																														
Further Evidence														x																
New Problems														x	x															
Application																														

@Agreement between teacher and criterion such by the observers. A close scrutiny of Table XXXIII suggests that the relatively large percentage (11.7%) assigned by teacher B to Identifying Problems may partly belong to Background Information. Similarly, the large percentage (16%) asigned by teacher C to New Problems may partly belong to Interpreting of Data.

It was considered reasonable to assume that if the formal observations showed a pattern similar to that of the long term summary, the long term patterns would be partly confirmed. To search for some regularity between the formal observations and the long term reports, the processes from the long term summary (Table XXXIII) were ranked according to the time spent on each process during the formal observation (see Table XXXVI). It was quite arbitrarily decided to reject all processes to which less than 6% of the process time was devoted. This figure was chosen simply from the consideration that, if each process occupied equal time, each would occupy approximately 6% of the total time.

Teachers D and E dealt with the same processes during the long term period. To a lesser extent, this was also the case for teacher A. Little similarity is evident between the long term summary and the formal observation for either teacher B or C except in the following instances:

TABLE XXXVI

PERCENTAGES AND RANKING OF TIME DEVOTED TO EACH PROCESS DURING LONG TERM PERIOD CONTRASTED TO PERCENTAGES OF TIME DURING FORMAL OBSERVATION

	LONG TERM PERCENTAGES FOLLOWED BY RANKING					FORMAL OBSERVATION PERCENTAGES				
	TEACHERS					TEACHERS				
	A	B	C	D	E	A	B	C	D	E
Identifying Prob.		11.7(4)		17.4(1)	6.7(6)				5.6	1.9
Background Info.	7.4(5)	24.2(1)	9.1(7)	9.9(5)	20.0(1)	8.2	62.9	7.1	4.7	9.8
Predictions	6.1(8)	10.8(5)	13.8(4)							
Hypotheses			10.5(6)					4.0		
Design	6.8(6)			15.0(3)	7.9(5)	13.5			5.6	3.9
Procedure	8.8(3)		13.1(5)	7.5(6)	12.5(4)	21.6			40.1	17.6
Observations	18.2(1)	10.4(6)	15.9(3)	17.0(2)	13.3(3)	21.6	29.0	49.5	30.8	42.2
Organizing Data	16.9(2)	13.3(3)			15.0(2)		6.4			3.9
Interpretations	6.8(7)	15.1(2)	16.8(1)	11.9(4)		8.2				
Further Evidence		6.7(7)								
New Problems			16.0(2)							
Applications	8.1(4)									

neither tended to emphasize Design, teacher B placed a great deal of emphasis upon Background Information, and teacher C placed some emphasis upon Hypotheses.

The tabulated results (Table XXXV) seem to contradict the essence of the process approach in the following: (1) teacher A placed very little emphasis upon Identification of Problems (0.5%) yet dealt extensively with Design (6.8%), Procedure (8.8%), and Observations (18.2%); (2) teacher B spent considerable time on Identification of Problems (11.2%), and Background Information (24.2%), and a minimal amount on Design (0.5%); (3) teacher C spent little time on Identifying Problems (0.5%), but a relatively large portion of the time on processes which normally follow the Identification of Problems.

Although no rigid order is implied in the Inventory within which framework the teachers under investigation were assumed to be operating, the design for the collection of data could scarcely be planned without a clear identification of the problem. Similarly, procedure and observations become almost random efforts unless structured by an appropriate design. These considerations suggest that teachers A, B, and C were not truly operating within the framework of the Inventory as would be interpreted by this researcher.

Coupled with the uncertainty thrown on whether teachers B and C perceived the processes in the same way as was for the criterion, it may be tentatively assumed that teachers B and C would rank slightly lower than A on a continuum ranging from "inconsistent" to "consistent" with the Inventory.

On the long term summary, teacher E reported that 40.8% of the time was devoted to Procedure, Observations, and Organizing Data and only 2.5% to Interpretations. This imbalance could perhaps occur in long term investigations for which interpretations had not yet been made but the investigations being carried out were of relatively short term duration. From this consideration, it was decided to rank E slightly below D in terms of consistency with the Inventory. In conformity with the foregoing arguments, the ranking of teachers in terms of consistency with the Inventory is depicted in Figure 5 below. It is not to be implied that this is an evaluative judgment in which D is in any way superior to any of the others but simply that D

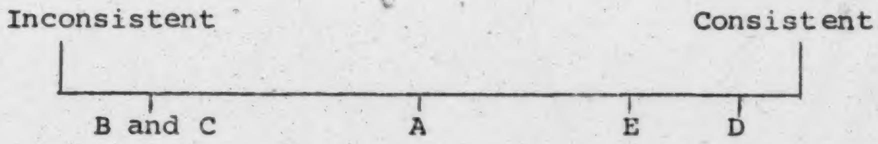


Figure 5. Ranking of Treatments in Terms of Consistency With the Inventory

appears to be the most consistent with respect to the Inventory and B and C the least consistent.

According to Powley (94), D was also superior in terms of attributes commonly associated with good teachers. The attributes, as listed by Powley, and the associated rankings based on the intuitive judgements of Powley and this researcher are presented in Appendix E.

A description of the different teaching acts is presented in Table XXXIX (Part II). Although this table samples only a small portion of the teaching time, some characteristics are worthy of note.

Teachers A and E were high on "teacher talks" and teachers C, D, and E low. Teachers E and C tended to ask questions of the factual recall type (35.6% and 37.5% respectively) as compared to 19.1% factual recall questions for teacher D and 10% for teacher A. Teachers D and A are thus more concerned with conceptual questions. Methods, with the exception of those of teacher B, were generally concrete as placed on a concrete--abstract continuum and the ranking of teachers on this continuum would be in the order A, E, D, C, and B. Subject matter was mainly theoretical (containing reference to abstract principles) as opposed to practical (totally within the experience of the learner). The teachers ranged from 64% theoretical to 100% theoretical and ranked in the increasing order C, E, A, B, and D. Because of the method of scoring used in the Observational Instrument, the term "theoretical" could be described as the "presence of theory in classroom activity". In the relatively abstract methodology of B, the presence of theory with classroom activity was high (98%). A more surprising result was the 100% presence of theory in the more concrete methodology of D.

The method used by each teacher during the formal observation partly confirms the information presented in Table XXXIV. All of the teachers except B used the laboratory as a major instructional mode. This table suggests that teachers D and E were more similar in their choice of instructional modes than were any other pair. The teachers are listed below followed in order by the instructional modes most often utilized by each.

A - Laboratory (40%), Discussion (24%), Field Study (10%)

B - Discussion (25%), A-V Media (23%), Library Research (12%)

C - Laboratory (32%), Discussion (19%), A-V Media (15%)

D - Laboratory (34%), Lecture (22%), Demonstration (15%), Discussion (13%)

E - Laboratory (36%), A-V Media (22%), Lecture (20%) Discussion (18%)

The characteristics of each teaching method as interpreted by this researcher from the available data are summarized in the following paragraphs:

Method A represented the extreme of the five examples in levels of permissiveness. Emphasis was placed upon active student participation. Theoretical concepts were in evidence and conceptual questions were favored. A minimal amount of time was spent on the identification of problems.

Method B was distinctly different from the other four. This methodology was the most abstract of the five. It was verbally oriented and student activity was vicarious rather than active. Scientific theory was dealt with extensively with little reference to activities in the laboratory or the field. There was a tendency to deal with the processes of science in isolation from each other.

Method C represented the extreme in structuring. Questions tended to emphasize factual recall, processes were dealt with as they arose from meticulously planned "investigations" and considerable emphasis was placed on supporting activities such as drill and review activities.

Methods D and E could be placed at the approximate mid-point on an unstructured-- structured continuum. Processes were dealt with as dictated by the content. The main discernible differences between these two methodologies were that D spent more time on supportive activities and his questions were more conceptually oriented.

Although it has been claimed that group D demonstrated superiority to the other groups in growth in knowledge about science and scientists as measured by the TOUS, it cannot be inferred that this was definitely due to any superiority in instructional procedures. Similarly, the superiority of D on the adjusted means of the PST and the IET may have been due to some initial innate superiority of group D. Although a cause-effect relationship cannot be established, some regularities are evident when treatment groups are ranked (as in Table XXXVII) in terms of the three categories of attitude, achievement, and characteristics of instructional input.

The attitude and achievement sections correspond closely but it is generally conceded that students with positive attitudes tend to do well and vice-versa. The

TABLE XXXVII

A COMPARISON OF THE RANKING OF THE GROUPS ON SELECTED INSTRUMENTS AND PROCEDURES

INSTRUMENT OR MEASURE		RANKING 1	2	3	4	5
TOUS	Achievement	D	E	C	B	A
SRT		D	E	B	C	A
PST		D	E	C	B	A
IET		D	C	A	B	E
HIFAMS (1)	Attitude	D	E	C	B	A
HIFAMS (2)		D	E	C	B	A
HIFAMS (3)		D	E	C	B	A
HIFAMS (4)		D	E	C	A	B
Consistency With the Inventory	Characteristics of Instructional Input	D	E	A	C	B
Ratio of Teacher-Student Talk to Teacher Talk		D	C	A	E	B
Ratio of Conceptual Question to Factual Questions		A	D	C	E	B
Nonverbal/verbal time ratio (laboratory+demonstration+audio-visual/total)		E	C	D	A	B

"Characteristics of Instructional Input" section demonstrate that, except for the position of A, the rankings under "Consistency With the Inventory" agree exactly with the rankings on the TOUS and the IET. The ranking of E is low on both the "Ratio of Teacher-Student Talk to Teacher Talk" and the "Ratio of Conceptual Questions to Factual Questions". The relatively low ranking of E and the relatively high ranking of A correspond to a similar relationship on the IET in which group A ranks third rather than its customary fifth and and group E ranks fifth rather than its usual second. There is little regularity, however, between the rankings of the three ratios (Teacher-Student Talk to Teacher Talk, Ratio of Conceptual Questions to Factual Questions, and Nonverbal/ Verbal Time Ratio) and the rankings in either the attitude or the achievement section.

The evidence suggests that the teaching strategy of "imitating the scientist" (Consistency With the Inventory) may be positively related to achievement in the process dimension. There is some indication that that a student's ability to select questions pertinent to the solution of a scientific investigation may be enhanced by teaching techniques which favor conceptual questions and thus encourage students' questions. There seems to be little support for any claim of a relationship between facets of an instructional procedure, such as the time ratios above, and student achievement in the process dimension.

IV. SUMMARY

In terms of the evaluation model of Chapter III (Figure 2), an attempt was made to determine the characteristics of each separate curriculum and relate these results to comparisons among groups in achievement in the pocess dimension.

It was determined that there were wide variations in presentations, implementations, and interpretations of the overall curriculum. Of the five methodologies observed, one was determined to be distinctly different from the other four. Of the remaining four methodologies, one was highly structured, one somewhat permissive, and the remaining two occupied a position midway between the two extremes.

There were significant differences among groups in knowledge about science and scientists, knowledge of and skill in the processes of science, and efficiency in inquiry procedures. These differences may have been due to some initial superiority in one of the groups but they could also be related to differences in instructional procedures.

There was little or no evidence to support any claim for a relationship between individual facets of an instructional procedure and student achievement in the process dimension. It is suggested that it is the combination and interrelationship of these facets which may be important.

Instructional procedures which tend to maximize process approach objectives appear to have the following characteristics: (a) the major instructional mode is a nonverbal (particularly laboratory) one, (b) the teacher asks conceptual questions and encourages students to ask further questions related to the phenomenon under investigation, and (c) the processes of science are dealt with as dictated by the subject matter under investigation. Further implications relating characteristics of input (treatment) to achievement will be found in the following chapter.

CHAPTER V

SUMMARY AND GENERALIZATIONS

I. SUMMARY

In this study an instructional model was designed and used to develop an evaluation model suitable for a process approach curriculum. The evaluation model was used to draw inferences relative to the problem of how to instruct students in science so that these students would gain knowledge of and skill in the processes of science and become more cognizant of the scientist and the scientific enterprise.

The evaluation model included the recognition that the intent and the implementation of a given curriculum are seldom the same and utilized observational procedures to estimate the actual nature of each curriculum used in the study. Evaluation instruments were selected, prepared, or adapted to sample behaviors considered to be indicative of proficiency in the process domain. These evaluation instruments purported to measure: (1) knowledge about science and scientists, (2) knowledge of and skill in the processes of science, and (3) efficiency in inquiry procedures.

A process approach biological science curriculum operating under the framework of the Inventory was the official curriculum used in the study. The actual curriculum was considered to consist of five separate curricula. The nature of each of these curricula was considered to depend upon how each participating teacher interpreted the official curriculum.

Comparisons were made among groups on achievement data and related to similarities and differences in input as estimated from the observational procedures in an attempt to determine the kinds of input which tended to maximize process approach objectives.

II. RESULTS

The results of the procedures outlined above are given in the previous chapter. These results indicate that the relative success of each group appears to be positively related to the degree to which the teachers understood and interpreted the processes of science in a manner consistent with the implications of the Inventory.

The recognition in the evaluation model that the intent and implementation of a given curriculum are seldom the same has been substantiated. The intent of the official curriculum was that each teacher would operate under the framework of the Inventory and thus accept the two major principles that (1) the subject matter is of primary importance and dictates the problem to be solved and the strategies of inquiry to be utilized and (2) students must be given the opportunity to consciously use the processes of scientific inquiry in learning science. The results of the observational procedures showed that two out of the five curricula were not truly operating within the framework of the Inventory and only two curricula were similar in levels of structuring and the use of the processes of scientific inquiry.

The instruments as selected, prepared, or adapted, were used to obtain data relative to an overall measure of success in understanding the process dimension of science by making comparisons among groups in areas specific to each instrument. The TOUS was used to make comparisons among groups in their knowledge about science and scientists. The Science Reasoning Test was used to measure knowledge of and skill in the processes of science. It dealt with the recognition of phases of a scientific investigation, the making of judgements related to the evaluation of hypotheses, and the ability to discern the suitability of particular experimental procedures for specific problems. The Process of Science Test was used to make comparisons related to how well students could actually operate in the process aspects of the curriculum when confronted with simulated investigations presented by visual stimuli. The Inquiry Efficiency Test utilized one half of each form of the TAB Science Test (57,19) but adopted a weighted scoring system to obtain a quite specific measure. This scoring system was used to obtain a measure related to the ability to select questions whose answers are helpful in obtaining a solution to a new or novel problem in science.

The two instruments specifically prepared for this study were the Science Reasoning Test and the Process of Science Test. Evidence about reliability and validity was presented in Chapter III. The Science Reasoning Test was found to be quite reliable but seemed to depend, to some extent, upon prior knowledge of science and tested intelligence. This was confirmed by the fact that no significant difference among treatment groups was found when the I,Q. scores and the COOP scores were used as covariates even although the difference was highly significant ($p < .001$) using analysis of variance. This test had been abbreviated so as to conform to the time available in regular classroom periods so there is some expectation that with further testing using an extended form both its reliability and sensitivity will be increased. The Process of Science Test used a prepared film loop to ensure that the stimuli would be the same for all groups tested. The film loop was supplemented by diagrams which emphasize key points of the investigation displayed by the film. Although lower in reliability than the Science Reasoning Test, it seemed to be less dependent upon prior knowledge and tested intelligence and more related to a performance ability.

III. DISCUSSION OF RESULTS

Although designed specifically as a measure of achievement in the process dimension, the technique used in the Process of Science Test could provide a link between evaluation and instruction. Actual classroom demonstrations or "investigations" could be videotaped and thus allow for student-teacher discussion of the suitability of the techniques and procedures employed. These videotaped recordings (or other suitable visual input such as film loops) could serve as problem foci for questions of three main types:

1. Questions dealing with the recognition of processes
2. Questions related to the ability to deal with specific processes
3. Questions which require the individual to exercise judgment about the suitability of the processes in relationship to the main problem.

Samples of questions of each of these types are to be found in the Science Reasoning Test and the Process of Science Test in Appendix C.

In process evaluation there would seem to be no adequate substitute for judgments related to actual performance. An evaluator of student competence in the process dimension must be in a position to answer the questions: To what degree does this student carry out an independent investigation?; Does he and can he hypothesize?, make accurate observations? and so on. It would be helpful if tests in the process dimension were supplemented by judgments made by teachers. It is suggested that these judgments be organized around the Inventory so that the teacher would grade each student's performance in each process used in an investigation. As an initial grading system, three grades could be assigned: yes (Y), no (N) and doubtful (D). On a sneet with the processes listed in the rows and the students' names in the columns, a quick recording is possible. This would provide an immediate visual reference for diagnostic purposes, not only for student grading and overall class competence, but also to illustrate whether the chosen investigations spanned the processes or simply concentrated on a few.

Although method D appeared to have been the most successful and method A (or possibly B) the least successful, a direct relationship between the characteristics of each teaching method and student achievement can only be implied on a descriptive basis.

Method A represented the extreme of the methodologies in levels of permissiveness and students were expected to discover many relationships with minimal assistance. The achievement of group A was low on all of the achievement tests except the IET. The unstructured inquiry approach, while not very successful in most areas, seems to be as successful as most of the others in increasing student ability to select questions whose answers are necessary to the solution of a scientific problem.

Method B was the most abstract and theoretical of the methodologies. Although group B was relatively low in most areas of achievement, it was approximately average on the SRT. A more abstract methodology may thus be slightly more effective than unstructured inquiry for dealing with more abstract data.

Method C was the most structured methodology. The investigations carried out by students were meticulously preplanned and the accompanying instructions were detailed and specific. The achievement of group C was slightly superior to that of group A or group B which could indicate that highly structured inquiry is more successful than unstructured inquiry at least at the junior high school level.

Methods D and E were quite similar and represent the approximate midpoint between A and C in terms of structuring. The main differences between the two methodologies were that D spent more time on supportive activities such as drill and review and his questions were more conceptually oriented. This observer made anecdotal notes during the observational sessions and had noted: "Students ask questions freely." The fact that group D was significantly superior to group E on the IET may reflect the differences between the two methodologies. Teacher E spent considerably more time than D in giving directions and also favored factual questions. The tendency of D to draw out questions from the students and direct them towards the answers rather than satisfying student queries by explicit directions may have a relationship to the superiority of group D.

As well as the continuum ranging from structured to unstructured opportunities for learning, there were differences among the methodologies in the choices of instructional modes and in the types of tasks posed for the students. If the tasks were beyond the intellectual or developmental level of the students, negative motivation could have resulted. On the other hand, if the tasks were simplistic and all of the materials were familiar, the incentive to discover new relationships could have been absent.

Waetjen (122) discusses epistemic motivation and indicates that motivation is increased if some element of either the content or instructional procedure is unknown or unpredicted by the student. If there is a perfect "match" between the curriculum content and instructional methods communicated to the individual's cognitive structure little learning takes place and similarly, if there is no "match" again little learning will result. This, of course, is what would be implied by Festinger's (33) theory of cognitive dissonance and could be illustrated by two extremes in science teaching strategies. At the one extreme a teacher could structure an investigation by means of specific instructions which detailed the procedures to be carried out and at the other extreme the teacher, in the name of inquiry, would present the student with an "investigation" entirely unfamiliar to the students and without any hints as to what the problem was all about or what procedures might be fruitful. Although no teacher in the group studied here belonged to either extreme, teacher C tended toward the first extreme and teacher A toward the second. In both instances the time devoted to the identification of problems was minimal, so motivation to learn may well have been slightly inhibited.

Many writers (115, 123, 16, 52) emphasize the importance of motivation and, directly or indirectly, the importance of closure. From the data in the previous chapter, it is evident that some members (A and B) of the teaching group seemed to minimize inferences resulting from investigations carried out by the student. The failure to complete the investigation circuit could conceivably have adversely affected student motivation and, hence, achievement.

The students used in this study were junior high school students and, in terms of development, have dependence on concrete empirical props. It is noted that treatment D included, as supplementary to the laboratory, a large amount of time devoted to demonstrations, Conversely, treatment B devoted almost no time to the laboratory and very little time to demonstrations. The relatively poor showing of group B as compared to D cannot be explained on this basis alone, but it is conceivable that treatment B might have been more suitable for senior high school students; i.e., the students in the sample were barely out of elementary school and could have had some difficulty dealing with the relatively more abstract treatment.

The failure to identify problems clearly and to allow students to flounder about in a relatively unstructured situation could cause discouragement and lead to a negative attitude. The success of a traditional approach in many instances may be associated with the fact that the direction and the approach the student was to use was set out explicitly. A study by Taylor Pearce (119) in the field of mathematics tends to substantiate the idea that creativity and inquiry are not increased by the provision of a relatively unstructured situation for investigation. This study found that a so-called "mathematizing" method which consisted of a stage of uninhibited exploration of a problem situation on the part of the pupils and a stage of hypotheses formulating by them was relatively inferior to the production of creativity than was the expository method. Strong evidence that the accumulation of knowledge was also related to, and possibly a prerequisite for, increased creativity adds support to the idea that substance and syntax of subject matter are interdependent.

The most successful of the treatments (D), as measured by the instruments used in this study, devoted almost forty percent (37.4%) of the time to lecture and demonstration. This could suggest that some expository treatment, skillfully done, is a very effective way of transmitting knowledge. When this is combined with investigations which form a cycle from the identification of the problem to the interpretations (which would provide closure and possibly increase motivation), further incentive to inquiry could be enhanced. It is entirely possible that the failure of some teachers to provide any answers and to concentrate entirely on process objectives may represent the swinging of the pendulum too far in the direction opposite to the subject matter approach. Meyer (75), in an evaluation of Nuffield Science, comments:

> The very frequent decline in interest shown by several Nuffield subjects in this study in solving problems by consulting authorative sources, such as books and journals, or by simply questioning the teacher is alarming. (75)

While skepticism definitely has its place, it would seem dangerous to do away with all authority figures and to encourage students at a very early age to reject commonly accepted scientific facts. Junior high school students, it would seem, do require some security, even in what they learn. There is a vast distinction between the notion of the teacher or written authority being sometimes wrong rather than always to be questioned. A great deal of skill is required to develop the idea that science searches after truth and its prime aim is to bring mankind closer to the truth. Schwab illustrates how the essential doubt component may be brought into its proper perspective:

> What is wanted, therefore, is a mode of discourse which will exhibit the hesitancies of inquiry for what they are: signs of the complexities of the problems which scientists dare solve, and objects of systematic research and rectification. (104)

As indicated previously, the Inventory was adopted in an attempt to provide a framework within which the product and the process of science could be shown as intimately related, neither as existing without the other. It was considered that the separation of the two within a teaching strategy would not reveal science as it is. Unfortunately, there is a distinct possibility that the Inventory was misinterpreted (two out of five of the teachers failed to agree with the criterion on the processes they were emphasizing). Over and above this, there is some possibility that the Inventory was misused to simply provide a list of science processes that the children could name and identify. This latter procedure could simply replace the old fashioned "Scientific Method" of some six categories with an equally formalized one containing seventeen categories and numerous sub-categories.

Although attitude was not found to be significantly related to science achievement within schools, there was a striking similarity between achievement and attitude scores. As a "post hoc" test, it was discovered that the attitude of students was significantly related to achievement (correlation coefficients between HIFAMS and TOUS posttest, SRT, PST and IET were 0.29 ($p < 0.001$), 0.29 ($p < 0.001$), 0.25 ($p < 0.001$) and 0.15 ($p < 0.01$) respectively). Although it could not be inferred that there was any causal relationship, it may be hypothesized that some part of the attitude score was directly related to instructional procedure.

Most individuals require some measure of success combined with some challenge and one of the most important tasks of the teacher could well be that of providing a balance between the success and the challenge. In dealing with the course within the Inventory, the challenge is presented with the identification of the problem and often the success with the resulting inference. It is the teacher's task to determine how much and what kind of assistance is required by each individual. Too much assistance might tend to stifle initiative and too little could result in withdrawal. It is not to be implied that all investigations must end with some definite inference, but any teacher who presents students with a series of seemingly everlasting tasks such as Problem → New Problem → New Problem, etc. is likely doomed to failure. It may not be entirely coincidental that the treatment groups (D and E) with the highest attitude scores devoted some time to new problems but did not overemphasize this process.

It is a truism that some teachers are better than others. In this particular study, no implication is made that any teacher was in any way superior but the methods used by some seemed to be more successful with regard to process teaching than the methods used by others. One difficulty inherent in this study was the superior quality of the participating teachers which tended to lessen the chance of discerning outstanding differences among achievement groups. Some tentative conclusions, not necessarily in the order of their importance, are now proffered:

1. Teachers of science should be familiar with the nature of science. This familiarity could be brought about by the teacher's thorough knowledge in some depth of a particular scientific discipline as well as knowledge obtained from selected readings about science.

2. Junior high school students should begin to plan and carry out their own investigations with the amount of dependence upon the teacher decreasing in relation to the student's growth in competence to work independently. Too much structuring, without a progressive decrease in the structuring, or unguided inquiry before the student feels competent to cope with the situation are the two extremes which should be avoided.

3. There seems to be no evidence that the most efficient ways of transmitting certain types of knowledge should

not be used to supplement the main instructional strategy of having the students carry out "investigations." (Teacher D seemed to have used lecture and demonstration methods to good advantage in increasing his students' competence in the process dimension.)

4. Although learning theory by itself is far from a sufficient guide for planning a science curriculum, classroom teachers should be aware of some of the generally accepted principles of learning theory. For example, instructional input should match the cognitive complexity and developmental level of the student, intrinsic motivation is important but it may be offset by social-ego pressures, and so on.

5. There is no evidence that an increase in the variety of instructional modes is related to the production of student competence in process knowledge and skill. The quality and appropriateness of a given instructional mode would seem to be more important.

6. The Inventory should be used primarily as a guide for the teachers. Its explicit use by the students should probably be incidental rather than formal. The knowledge of the processes may be best developed as a result of active participation by the students in investigations involving

the particular processes. A knowledge of the gross processes as a sort of inquiry cycle may be essential.

7. Process evaluation should be carried out using tests which contain problem foci closely akin to actual investigations. It is recommended that these tests be supplemented by teacher judgments directly related to the instructional procedure. In the interests of objectivity, judgments should be made in accordance with a pre-planned observational schedule containing a list of the process skills and abilities being evaluated and gradings should be recorded in no more than three categories: competent, not competent and doubtful. This close alliance of teaching and testing should clarify the objectives for both the teacher and the student.

8. The teacher factor rather than the name of the official curriculum determines what curriculum is in operation in each classroom. The intent of a curriculum and the practice of a teacher are often dissimilar.

II. FOR FURTHER RESEARCH

In this study an evaluation of achievement in the process dimension was attempted in conjunction with observational procedures in order to gain further information about the techniques and methodologies which would maximize the attainment of process approach objectives.

A replication of the present study could be carried out by a team of researchers intimately associated with the administrative and instructional staff of a large school division. A team approach could not only determine the time devoted to each instructional mode and to each process but also the timing and sequencing of the instructional modes as related to the substantive content and the associated processes of scientific inquiry. In this manner a more comprehensive picture of the characteristics of instructional input could be obtained. Futher, the close association of the research with the instructional procedures could make time available for the inclusion of more instruments (including pretesting) in order to ascertain gains in all areas of achievement.

As well as these possible refinements to the present study, the following associated considerations warrant further study:

(a) The two methodologies which appeared to be the most successful were neither highly structured or relatively unstructured. Further study is needed in an attempt to determine the optimum structuring in relation to developmental stages.

(b) The close correspondence between attitude and achievement rankings suggest that growth in attaining the affective attibutes of scientists should not remain a broad unmeasured objective. Nay and Crocker (81) have prepared a comprehensive list of these attributes and Flynn and Munroe (36) have demonstrated that some of the attributes associated with the intellectual performance of scientists can be measured behaviorally. There is need for more instruments for the assessment of growth in this area and the subsequent application of the findings of these instruments to the development of instructional strategies.

(c) Each of the disciplines in science requires certain specific skills and information of those who would achieve any degree of competence in the discipline. The knowledge of some minimum number of chemical symbols and the ability to deal with physical units of measurements are examples. Comparative studies could contrast the achievement of groups who used S-R models as an efficient way of gaining necessary knowledge with the achievement of groups who gained the information incidentally. Studies of this kind could determine the place of S-R models in an overall instructional model for the specific disciplines.

(d) Too little is known about the matching of students with instructional procedures. A study using a multivariate approach with aptitude tests, attitude scales, interest inventories, cognitive preference tests, and the usual battery of achievement tests associated with a firm observational schedule might possibly isolate the relevant variables.

(e) It was noted in this study that the most successful of the teachers (D) appeared to link theoretical concepts to classroom activities and was particularly skilful in drawing out questions from his students. Studies are needed to explore the phenomenon which may be referred to as "linkage". What techniques link theoretical concepts to classroom activities and to the cognitive functioning of the students? Do these techniques form a pattern which may be learned and then adapted to the teaching strategy of another?

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Aiken, Lewis R. "Nonintellective Variables and Mathematics Achievement Directions for Research," Journal of School Psychology, Vol. 8, No. 1, 1970, pp. 28-35.

2. Aikenhead, Glen S. "The Measurement of High School Students' Knowledge About Science and Scientists." Unpublished Qualifying Paper, Harvard Graduate School of Education, 1970.

3. American Association for the Advancement of Science, The Process Instrument, (AAAS Miscellaneous Publication 65-25), 1965.

4. American Association for the Advancement of Science, Science a Process Approach, Commentary for Teachers, 3rd ed., (AAAS Miscellaneous Publication 68-7), 1968.

5. Anastasi, Anne (ed.). Testing Problems in Perspective. Princeton: Educational Testing Service, 1965.

6. Arons, Arnold. "Toward Wider Public Understanding of Science," American Journal of Physics, Vol. 41/6, June, 1973, pp. 769-782.

7. Ausubel, David P. Educational Psychology A Cognitive View. New York: Holt, Rinehart and Winston, Inc., 1968.

8. Ausubel, David P. "Some Psychological and Educational Limitations of Learning by Discovery," The Arithmetic Teacher, May, 1964, pp. 290-302.

9. Ausubel, David P. The Psychology of Meaningful Verbal Learning. New York: Grune and Stratton, 1963.

10. Biddle, Bruce J. "Methods and Concepts in Classroom Research," Review of Educational Research, XXXVII(3), 1967, pp. 337-357.

11. Biological Science Curriculum Study. Process of Science Test, Form A. New York: The Psychological Corporation, 1965.

12. Bloom, Benjamin S. "Learning for Mastery," Evaluation Comment, May, 1968.

13. Bloom, Benjamin S. (ed.). Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: David McKay Co. Inc., 1956.

14. Bruner, Jerome S. "The Act of Discovery," Harvard Educational Review, Vol. 31, No. 1, 1961, pp. 21-32.

15. Bruner, Jerome S. The Process of Education. Cambridge: Harvard University Press, 1962.

16. Bruner, Jerome S. Toward a Theory of Instruction. Cambridge: Belknap Press of Harvard University, 1966.

17. Burmester, Mary Alice. "The Construction and Validation of a Test to Measure Some of the Inductive Aspects of Scientific Thinking," Science Education, XXXVII, March, 1953, pp. 131-144.

18. Buros, O.K. (ed.). The Seventh Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon Press, 1963, pp. 1240-1243.

19. Butts, David P. and Howard L. Jones. "The Development of the Tab Science Test,""Science Education, LI, December, 1967, pp. 463-374.

20. Carroll, John B. "A Model of School Learning," Teachers' College Record, LXIV, pp. 723-733.

21. Cattell, Raymond B. (ed.). Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally and Co., 1966.

22. Cooley, William W. and Robert D. Bassett. "Evaluation and Follow-up Study of a Summer Science and Mathematics Program for Talented Secondary School Students," *Science Education*, XLIV, April, 1961, pp. 209-216.

23. Cooley, William W. and Leo Klopfer. *Test on Understanding Science: Form W*. Princeton, New Jersey: Educational Testing Service, 1961.

24. Cooperative Science Tests. *General Science Form B*. Princeton, New Jersey: Educational Testing Service, 1962.

25. Cooperative Science Tests. *Cooperative Science Tests Handbook*. Princeton, New Jersey: Educational Testing Service, 1964.

26. Cossman, George W. "The Effects of a Course in Science and Culture for Secondary School Students," *Journal of Research in Science Teaching*, VI, September, 1969, pp. 274-283.

27. Crocker, Robert K. "The Suchman Project on Elementary School Science Inquiry," Edmonton: Department of Secondary Education, University of Alberta, 1967. (Mimeographed.)

28. Crumb, Glenn H. "Understanding of Science in High School Physics," *Journal of Research in Science Teaching*, III(3), 1965, pp. 246-250.

29. Dewey, John. *The School and Society*. Chicago: University of Chicago Press, 1900.

30. Draper, N. R. and H. Smith. *Applied Regression Analysis*. New York: John Wiley and Sons, 1966, pp. 178-194.

31. Edmonton Junior High School Life Science Committee. "Life Science a Process Approach," (Chapters 1 and 2). Edmonton: Edmonton Junior High School Life Science Committee, 1969. (Mimeographed.)

32. Ferguson, George A. Statistical Analysis in Psychology and Education. New York: McGraw Hill, 1966.

33. Festinger, Leon A. A Theory of Cognitive Dissonance. Evanston: Row, Peterson, 1957.

34. Fischler, Abraham S. and George Zimmer. "The Development of an Observational Instrument for Science Teaching," Journal of Research in Science Teaching, V(2), 1967-1968, pp. 127-137.

35. Flanders, Ned A. "Interaction Analysis and Inservice Training," Journal of Experimental Education, XXXVII, pp. 127-133.

36. Flynn, H.E. and R.G. Munro. "An Evaluation of Nuffield Science," The School Science Review, L. November, 1968, pp. 394-402.

37. Friot, Faith E. "The Relationship Between an Inquiry Teaching Approach and Intellectual Development," Unpublished Doctoral Dissertation, University of Oklahoma, 1970.

38. Gage, N.L. "Successful Teacher Behavior: Can Science Contribute to the Art of Teaching?" Paper read at the Proceedings of the Second International Conference on Elementary Education, Banff, Alberta, October 25-28, 1967.

39. Gage, N.L. and W.R. Unruh. "Theoretical Formulations for Research on Teaching," Review of Educational Research, XXXVII (3), 1967, pp. 359-370.

40. Gagné, Robert M. The Conditions of Learning. New York: Holt Rinehart and Winston Inc., 1967.

41. ________. "Military Service and Principles of Learning," American Psychologist, XVII, February, 1962, pp. 83-91.

42. ________. "The Psychological Bases of *Science-A Process Approach*," American Association for the Advancement of Science, (AAAS Miscellaneous Publication), 1965.

43. Gallagher, James J. "A 'Topic Classification System,' for Classroom Interaction," *Classroom Observations*, AERA Monograph Series on Curriculum Evaluation, No. 6. Chicago: Rand McNally and Company, 1970.

44. Glaser, Robert F. "Concept Learning and Concept Teaching," in *Learning Research and School Subjects* (R. Gagne and W.J. Gephant (eds.)), Pittsburgh: F.E. Peacock Publishers, 1968, pp. 1-38.

45. Grobman, Hulda. "The Rationale and Framework of the BSCS Evaluation Program," *BSCS Newsletter No. 19*. Boulder, Colorado: Biological Science Curriculum Study, September, 1963.

46. ________. "Background of the 1963-1964 Evaluation," *BSCS Newsletter No. 24*. Boulder, Colorado: Biological Science Curriculum Study, September, 1963.

47. ________. *Evaluation Activities of Curriculum Projects: A Starting Point*, AERA Monograph Series on Curriculum Evaluation No. 2. Chicago: Rand McNally and Company, 1968.

48. Guilford, J.P. *The Nature of Human Intelligence*. New York: McGraw-Hill Book Company, 1967, pp. 312-345.

49. Gulliksen, Harold. *Theory of Mental Tests*. New York: John Wiley and Sons Inc., 1950, pp. 11-27.

50. Harman, Harry H. *Modern Factor Analysis*. Chicago: The University of Chicago Press, 1967.

51. Hedley, Robert Lloyd. "Student Attitude and Achievement in Manitoba Secondary Schools." Unpublished Doctoral Dissertation, Michigan State University, East Lansing, 1966.

52. Holt, John. How Children Fail. New York: Pitman Publishing Company, 1964.

53. Hubbard, J.P. "An Objective Evaluation of Clinical Competence," The New England Journal of Medicine, 272:1321-1328, 1965.

54. Hukins, Austin A. "A Factorial Investigation of Measurements of Achievements of Objectives in Science Teaching." Unpublished Doctoral Dissertation, University of Alberta, 1963.

55. Inhelder, Barbel and Jean Piaget. The Growth of Logical Thinking From Childhood to Adolescence. New York: Harper and Row, 1961.

56. Jackson, Philip W. "The Way Teaching Is," in, The Way Teaching Is. Washington: ASCD/NEA, 1966. pp. 7-27.

57. Jones, Howard Leon. "The Development of a Test of Scientific Inquiry Using the TAB Format and an Analysis of Its Relationship to Selected Student Behaviors and Abilities." Unpublished Doctoral Dissertation, The University of Texas, January, 1966.

58. Kaiser, Henry F. and John Caffrey. "Alpha Factor Analysis," Psychometrika, XXX(1), March, 1965, pp. 1-14.

59. Kaplan, Eugene H. "The Burmester Test of Aspects of Scientific Thinking as a Means of Teaching the Mechanics of the Scientific Method," Science Education, LI(4), October, 1967, pp. 354-357.

60. Keeping, E.S. Introduction to Statistical Inference. New York: Van Nostrand, 1962, pp. 214.

61. Klopfer, Leopold E. and Donald C. McCann. "Evaluation of Unified Science: Measuring the Effectiveness of the National Science Course at the University of Chicago High School," Science Education, LIII(3), March, 1969.

62. Kratwohl, David R. et al. Taxonomy of Educational Objectives Handbook II: Affective Domain. New York: David McKay Company Inc., 1964.

63. Kruglak, Haym. "The Measurement of Laboratory Achievement," American Journal of Physics, XXII, 1954, pp. 442-462.

64. ________. "The Measurement of Laboratory Achievement, Part III - Paper-Pencil Analogs of Laboratory Performance Tests," The American Journal of Physics, XXIII, 1955, pp. 82-87.

65. ________. "Some Behavior Objectives for Laboratory Instruction," American Journal of Physics, XIX, 1951, pp. 223-225.

66. ________. "Resource Letter AT-1 on Achievement Testing," American Journal of Physics, XXIII, 1965, pp. 255-263.

67. Mallinson, George G. in O.K. Buros, (ed.). The Seventh Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon Press, 1963, pp. 1227-1228.

68. Mayer, William V. in J. David Lockard (ed.). "Biological Science Curriculum Study (BSCS)," Seventh Report of the International Clearinghouse on Science and Mathematics Curricular Developments. A joint project of the American Association for the Advancement of Science and the Science Teaching Center, University of Maryland, College Park, Maryland, 1970, pp. 277-282.

69. Maykovitch, John J. "Relationships Between Attitudinal Orientations and Academic Achievement in the Secondary School." Unpublished Doctoral Dissertation, University of California, Berkley, 1966.

70. McKinnon, Joe W. and John W. Renner. "Are Colleges Concerned With Intellectual Development?" American Journal of Physics, XXXIX, September, 1971, pp. 1047-1052.

71. Medley, Donald M. and Harold E. Mitzel. "Measuring Classroom Behavior by Systematic Observation," in N.L. Gage (ed.) Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963, pp. 247-327.

72. ________, ________. "A Technique for Measuring Classroom Behavior," Journal of Experimental Psychology, XLIX(2), 1958, pp. 87-92.

73. ________, ________. "Some Behavior Correlates of Teacher Effectiveness," The Journal of Educational Psychology, L(6), 1959, pp. 239-246.

74. ________, and R.A. Hill. "A Comparison of Two Techniques for Analyzing Classroom Behaviors." Paper delivered at AERA Annual Meeting, 1968.

75. Meyer, G.R. "Reaction of Pupils to Nuffield Science Teaching Project Trial Materials in England at the Ordinary Level of the General Certificate of Education," Journal of Research in Science Teaching, VII(4), pp. 283-302.

76. Mokosch, Eric. "The Development and Evaluation of a Process Approach to the Teaching of Junior High School Science." Unpublished Doctoral Dissertation, University of Alberta, Edmonton, 1969.

77. Morrison, Donald F. Multivariate Statistical Methods. New York: McGraw-Hill Book Company, 1967, pp. 117-154.

78. Nay, Marshall A. "Behavioral Objectives on the Process of Science." SCAT Bulletin, Vol. IX, Numbers 2 and 3, pp. 77-81, September 1970.

79. ________. "The Discovery Method of Teaching," The ATA Magazine, Vol. 48, No. 5, pp. 20-23, March - April, 1968. Reprinted in abridged form in Teaching for Tomorrow: A Symposium on the Social Studies in Canada, J. Lewis (ed.), Toronto: Thomas Nelson & Sons, 1969.

80. ________, and Associates. "A Process Approach to Science Teaching," Science Education, LV(2), 1971, pp. 197-207.

81. ________, and Robert K. Crocker. "Science Teaching and the Affective Attributes of Scientists," Science Education, LIV(1), 1970, pp. 59-67.

82. Newport, John F. and Keith McNeil. "A Comparison of Teacher-Pupil Verbal Behavior Evoked by Science - A Process Approach and by Textbooks," Journal of Research in Science Teaching, VII(3), 1970, pp. 191-195.

83. Newton, David E. "The Dishonesty of Inquiry Teaching," School Science and Mathematics, LXVIII, 1968, pp. 807-810.

84. Novak, Joseph D. "A Case Study of Curriculum Change - Science Since PSSC." Paper presented at the conference on, "Curriculum Development in a Changing World," Syracuse University, July 10-11, 1968.

85. Nuthall, Graham A. "An Experimental Comparison of Alternative Strategies for Teaching Concepts," American Educational Research Journal, V(4), November, 1968.

86. ________. "A Review of Some of the Selected Recent Studies of Classroom Interaction and Teaching Behavior," Classroom Observation. AERA Monograph Series on Curriculum Evaluation, No. 6. Chicago: Rand McNally, 1970.

87. Okey, James R. and Robert M. Gagne. "Revision of a Science Topic Using Evidence of Performance on Subordinate Skills," Journal of Research in Science Teaching, VII(4), 1970, pp. 321-325.

88. Parker, J. Cecil and Louis J. Rubin. Process as Content. Chicago: Rand McNally and Company, 1968.

89. Pella, Milton O. "The Laboratory and Science Teaching," The Science Teacher, September, 1961, pp. 30-31.

90. Piaget, Jean. The Child's Conception of Physical Causality. New York: Harcourt, Brace and Company, 1930.

91. Pimental, George C. in David J. Lockard (ed.). "Chemical Education Materials Study (CHEMS)," Seventh Report of the International Clearinghouse on Science and Mathematics Curricular Developments. A joint product of the American Association for the Advancement of Science and the Science Teaching Center, University of Maryland, College Park, Maryland, 1970, pp. 303-306.

92. ________, in Preface to Chemistry an Experimental Science. San Fransisco: W.H. Freeman and Company, 1963.

93. Powley, David P. in Preface to "Life Science a Process Approach." Edmonton: The Life Science Committee, 1969. (Mimeographed.)

94. ________. "Teaching for Process in Scientific Inquiry." Unpublished Master's Dissertation, University of Alberta, Edmonton, 1970.

95. Programme of Studies for the Schools of Manitoba, Grades VII to XII. Winnipeg: Manitoba Department of Education, 1931, pp. 78-79.

96. Pyra, Joseph Frederick. "A Study of the Relationship Between School Characteristics and Student Attitudes Toward the School." Unpublished Master's Thesis, University of Alberta, Edmonton, 1965.

97. Ramsey, Gregor A. and Robert W. Howe. "An Analysis of Research in Instructional Procedures in Secondary School Science," The Science Teacher, XXXVI, pp. 62-81.

98. Rathes, James et al. *Studying Teaching*. Englewood, Cliffs, New Jersey: Prentice Hall Inc., 1967.

99. Renne, Thomas, Heidi Kass and Marshall A. Nay. "The Effect of Verbalizers on the Achievement of Non-Verbalizers in an Enquiring Classroom," *Journal of Research in Science Teaching*, X(2), 1973, pp. 113-124.

100. Raun, Chester E. and David P. Butts. "The Relationship Between the Strategies of Inquiry in Science and Student Cognitive and Affective Behavioral Change," *Journal of Research in Science Teaching*, V, 1967-68, pp. 261-268.

101. Rothman, Arthur I., Wayne W. Welch and Herbert J. Walberg. "Physics Teachers' Characteristics and Student Learning," *The Journal of Research in Science Teaching*, VI, 1969, pp. 69-73.

102. Schwab, Joseph J. "Structure of the Disciplines: Meanings and Significances," in G.W. Ford and Lawrence Pugno, (eds.), *The Structure of Knowledge and the Curriculum.* Chicago: Rand McNally and Company, 1964, pp. 6-30.

103. ________. "The Structure of the Natural Sciences," in G.W. Ford and Lawrence Pugno, (eds.), *The Structure of Knowledge and the Curriculum*. Chicago: Rand McNally and Company, 1964, pp. 31-49.

104. ________, and Paul F. Brandwein. *The Teaching of Science*. Cambridge: Harvard University Press, 1966.

105. Scientific Literacy Center. *Wisconsin Inventory of Science Processes*. Wisconsin: The Regents of the University of Wisconsin, 1967.

106. ________. "Key to WISP," Madison, Wisconsin, 1967. (Mimeographed.)

107. Siegel, Lawrence (ed.). *Instruction, Some Contemporary Viewpoints*. San Fransisco: Chandler Publishing Company, 1967, pp. 261-290.

108. Smith, Eugene R. and Ralph W. Tyler. Appraising and Recording Student Progress. New York: Harper and Brothers, 1942.

109. Smith, B. Othaniel et al. A Tentative Report on the Strategies of Teaching. Urbana: University of Illinois, 1964.

110. Stake, Robert E. "The Decision: Does Classroom Observation Belong in an Evaluation Plan?" Classroom Observation, AERA Monograph Series on Curriculum Evaluation, No. 6, Chicago: Rand McNally and Company, 1970, pp. 1-5.

111. Stephens, J.M. The Process of Schooling. New York: Holt, Rinehart and Winston, 1967.

112. Stice, Glen. Facts About Science Test. Princeton, New Jersey: Educational Testing Service, 1958.

113. ________. Manual for the Facts About Science Test. Princeton, New Jersey: Educational Testing Service, 1958.

114. Stotler, Donald W. "Science Heuristics and Humanism," The Science Teacher, XXXII, October, 1965, pp. 28-31.

115. Suchman, J. Richard. The Elementary School Training Program in Scientific Inquiry. Urbana: University of Illinois, 1962.

116. ________. "Inquiry," The Instructor, August, 1966-July, 1967. (A Model for the Language of Education," August - September, 1966; "Action Control and Intake," October, 1966; "Functions of Storage," November, 1966; "Motivation," December, 1966; "Old Goals and New Perspectives," January, 1967; "Play and Discussion," February, 1967; "Didactics," March, 1967; "Diagnostic Teaching," April, 1967; "Three Old Standbys," May, 1967; "Todays Problems - Tomorrows Possibilities," June - July, 1967.)

117. Tanner, Daniel. "Curriculum Theory, Knowledge and Content," Review of Educational Research, XXXVI, June, 1966, pp. 362-372.

118. Tanner, R. Thomas. "Expository-Deductive Versus Discovery-Inductive Programming of Physical Science Principles," Journal of Research in Science Teaching, VI, 1969, pp. 136-142.

119. Taylor-Pearce, J. Modupe. "A Study of the Relative Effectiveness of Two Teaching Methods with Respect to (Divergent Thinking) Creativity in Mathematics at the Grade Eleven Level." Unpublished Master's Dissertation, University of Alberta, Edmonton, 1969.

120. Thorndike, Robert L. and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1955.

121. Trent, John. "The Attainment of the Concept 'Understanding Science' Using Contrasting Physics Courses," Journal of Research in Science Teaching, III(3), 1965, pp. 224-229.

122. Waetjen, Walter B. "Learning and Motivation: Implications for the Teaching of Science," The Science Teacher, May, 1965, pp. 22-25.

123. Wall, Clifford N., H. Kruglak, and L.E.H. Trainor. "Laboratory Performance Tests at the University of Minnesota," American Journal of Physics, XIX, 1951, pp. 546-551.

124. Welch, Wayne W. "Curriculum Evaluation," Review of Educational Research, XXXIX, October, 1969, pp. 429-443.

125. ________, and Milton O. Pella. "The Development of an Instrument for Inventorying Knowledge of the Processes of Science," Journal of Research in Science Teaching, V, 1968, pp. 64-68.

126. Welch Science Process Inventory, Form D. Minneapolis, Minn.: Dr. Wayne W. Welch, 1969.

127. ________, and Arthur I. Rothman. "The Success of Recruited Students in a New Physics Course," Science Education, April, 1968, pp. 270-273.

128. Wheeler, Sherman E. "Critique and Revision of an Evaluation Instrument to Measure Students' Understanding of Science and Scientists." Chicago: University of Chicago, 1968. (Mimeographed.)

129. Wilson Brother Ralph C.S.C. "The Grading of Laboratory Performance in Biology Courses," in Hans O. Andersen, Readings in Science Education for the Secondary School. New York: The Macmillan Company, 1969, pp. 169-174.

130. Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

131. Yager, Robert E. "Teacher Effects on the Outcomes of Science Instruction," Journal of Research in Science Teaching, IV, 1966, pp. 236-242.

APPENDIX A

An Inventory of Processes in Scientific Inquiry

AN INVENTORY OF PROCESSES IN SCIENTIFIC INQUIRY

(Revised Edition*, 1970, as submitted to Science Education)

THE SPECIFIC PROCESSES

I. Initiation

1. Identifying and formulating a problem
    - (a) speculating about a phenomenon
    - (b) identifying variables
    - (c) noting and making assumptions
    - (d) delimiting the problem
2. Seeking relevant background information
    - (a) recalling relevant knowledge and experiences
    - (b) doing literature research
    - (c) consulting people
3. Predicting
4. Hypothesizing
5. Design for collection of data through field work and/or experimentation
    - (a) defining the independent and control variables in operational terms.

---

*The Fifth Draft of the Inventory, August 1968 appears in a Doctoral Dissertation by E. Mokosch (76).

    (b) defining the procedure and sequencing the steps
    (c) identifying needed equipment, materials and techniques
    (d) indicating safety precautions
    (e) devising the method for recording data

II. Collection of Data

6. Procedure
    (a) collecting, constructing, and setting up the apparatus or equipment
    (b) doing field work and/or performing the experiment
    (c) identifying the limitations of the design (as a result of failures, blind alleys, etc.) and modifying the procedure (often by trial-and-error)
    (d) repeating the experiment (for reproducibility, to overcome limitations of initial design, etc.)
    (e) recording data (describing, tabulating, diagramming, photographing, etc.)

7. Observing and observations
    (a) obtaining qualitative data (using senses, etc.)
    (b) obtaining semi-quantitative and quantitative data (measuring, reading scales, calibrating,

counting objects, or events, estimating, approximating, etc.)

- (c) gathering specimens
- (d) obtaining graphical data (charts, photographs, films, etc.)
- (e) noting unexpected or accidental occurrences (serendipity)
- (f) noting the precision and accuracy of data
- (g) judging the reliability,and validity of data

III. Processing of Data

- 8. Organizing the data
  - (a) ordering to identify regularities
  - (b) classifying
  - (c) comparing
- 9. Representing the data graphically
  - (a) drawing graphs, charts, maps, diagrams, etc.
  - (b) interpolating, extrapolating, etc.
- 10. Treating the data mathematically
  - (a) computing (calculating)
  - (b) using statistics
  - (c) determining the uncertainty in the results

IV. Conceptualization of Data

- 11. Interpreting the data

(a) suggesting an explanation for a set of data

(b) deriving an inference or generalization from a set of data

(c) assessing validity of initial assumptions, predictions and hypotheses

12. Formulating operational definitions

(a) verbal

(b) mathematical

13. Expressing data in the form of a mathematical relationship

14. Incorporating the new discovery into the existing theory (developing a "mental model")

V. Open-endedness

15. Seeking further evidence to

(a) increase the level of confidence in the explanation or generalization

(b) test the range or applicability of the explanation or generalization

16. Identifying new problems for investigation because of

(a) the need to study the effect of a new variable

(b) anomalous or unexpected observations

(c) incompleteness ("gaps") and inconsistence in the theory

17. Applying the discovered knowledge

APPENDIX B

Teacher Report Formats and Instructions to Teachers

INSTRUCTIONS FOR THE RANKING OF STUDENTS ON THE BASIS OF ATTITUDE

SORT EACH GROUP OF CARDS SO THAT THE NAME OF THE PERSON WITH THE BEST ATTITUDE TOWARD SCHOOL (NOT NECESSARILY SCIENCE ONLY) IS PLACED AT THE TOP OF THE PILE, THE NAME OF THE SECOND BEST IN THE NEXT POSITION AND SO ON. SOME JUDGMENTS MAY BE DIFFICULT, SO IN CASES OF DOUBT YOU ARE ASKED TO ARBITRARILY MAKE A SELECTION. THESE SELECTIONS SHOULD BE MADE FOR EACH SMALL GROUP OF SIX AS GIVEN. DO NOT ATTEMPT TO GO BACK AFTER YOU HAVE MADE YOUR INITIAL SELECTION.

ATTITUDE IN THE SENSE USED HERE IS SIMPLY YOUR (THE TEACHER'S) CONCEPTION OF A GOOD POSITIVE ATTITUDE TOWARD SCHOOL.

IT IS EXTREMELY IMPORTANT THAT THE GROUPS REMAIN AS GIVEN. SUGGESTION: REMOVE PAPER CLIP OR RUBBER BAND FROM ONE GROUP, ARRANGE THESE IN ORDER AND CLIP THEM TOGETHER BEFORE LOOKING AT THE NEXT GROUP. THIS MAY BE DONE QUITE RAPIDLY AND THEN QUICKLY CHECKED TO MAKE SURE THAT THE CARDS ARE PLACED IN THE ORDER OF YOUR CHOICE.

REPORT/: Teaching for Process in Scientific Inquiry

SCHOOL/Name__________ Date__________ Class Time__________

LESSON (Concepts/Content)

(General description of situation)

INQUIRY MODES Time (Fraction)

Laboratory ____ ____
Field Trip ____ ____
Invitations to Inquiry ____ ____
Lecture ____ ____
Case Study ____ ____
Simulation ____ ____
Demonstration ____ ____
A.V. ____ ____
Discussion ____ ____
Project ____ ____
Other (state) ____ ____

SCIENCE PROCESSES (Behavior Objectives) (Check off)

Preparation
1. Ident. & Form. of problem ____
2. Background Inf. ____
3. Predictions ____
4. Hypothesis ____
5. Design ____

Experimentation
6. Procedure ____
7. Observation ____

Processing of Data
8. Organizing data ____

Processing of Data (cont'd)
9. Graphing ____
10. Mathematical treatment ____

Conceptualization of Data
11. Operational definitions ____
12. Inference ____
13. Mathematical relationship ____
14. Mental models ____

Open Endedness
15. Further evidence ____
16. New Problems ____
17. Application ____

INSTRUCTIONAL PROCEDURE (Teaching for Process)

(a) How was "process" incorporated into the lesson, i.e., made operational?
-what did you as a teacher do?
-how was the class motivated?
-how were associated and basic skills taught?
-what materials and methods were used?
-what were the problems and your solutions?
-be specific, i.e., questions

(cont'd)

INSTRUCTIONAL PROCEDURE (continued)
(Teaching for Process)

asked, presentations to pupils, etc. Indicate time (min.)

(b) Evaluate the effectiveness of what was done and indicate what modifications you would make.

APPENDIX C

Tests and Instruments

TEST ON UNDERSTANDING SCIENCE FORM Ew

DIRECTIONS

1. Each of the questions or incomplete statements in this booklet is followed by four possible answers. You are to choose the one BEST answer for each question.

2. All of your answers are to be made on the separate answer sheet you have been given. Do not make any marks in this booklet.

3. Indicate your answers on the answer sheet by completely BLACKENING the box which has the same capital letter as the answer you have chosen. You may use a regular pencil to make your marks. However, do not use a fountain pen, ball point pen, or colored pencil.

4. You are to mark one, and only one, answer for each question. If you change your mind about an answer, erase it completely and cleanly. Be sure to make your new mark heavy and dark.

Here is an example:

0. The main reason for doing experiments in science is to

A. check ideas.
B. find things out.
C. use equipment.
D. learn about nature

ANSWER SHEET

	A	B	C	D
0.	☐	☐	■	☐

Imagine that the right half of the above example is a small piece of your separate answer sheet. Look

(continued)

at row "0" on this "midget" answer sheet. In the example, answer box C has been blackened. This means that the person thinks that the main reason for doing experiments in science is to use equipment. Your answers for all the questions in this booklet should be marked in this manner on your separate answer sheet.

6. Be sure that the number of the row you are marking and the number of the question you are answering are the SAME. If you have any questions about these directions, raise your hand. Your teacher will try to help you.
7. Try hard to answer all of the questions in this booklet.
8. Once again, DO NOT WRITE ANYTHING IN THIS QUESTION BOOKLET.

Please do not turn this page until you are asked to do so.

1. Let's compare scientists and magicians. Which one of the following sentences is best?

   A. Scientists and magicians both try to explain things.
   B. Scientists and magicians both try to make things mysterious.
   C. Magicians try to explain things, but scientists make them mysterious.
   D. Scientists try to explain things, but magicians make them mysterious.

2. Scientific discoveries have come from

   A. almost all countries of the world.
   B. only countries with big industries.
   C. only countries with large populations.
   D. almost all countries in Africa and Asia.

3. George said: "A scientist's work never ends." By this, George means that

   A. the work day in a laboratory is very long.
   B. people ask scientists many difficult questions.
   C. scientists need more help in their research.
   D. new questions come up whenever a problem is solved.

4. When a scientist finishes a research project, he will usually

   A. keep his results secret to help his country.
   B. let other scientists know about his findings and ideas.
   C. use his results in an invention for industry.
   D. ask the government for permission to write a report.

5. Scientists today agree on many ideas about how the natural world works. Most likely, these ideas will

   A. be changed when scientists have more information.
   B. not be changed for a very long time to come.

(continued)

5. (continued)

C. be changed to keep up with fast-moving world events.
D. not be changed because they are scientific ideas.

6. Today, the education of American scientists who teach and do research at colleges and universities generally

A. is completed after four years of college attendance.
B. includes a period of practical training in some industry.
C. is completed after two years of college or technical school.
D. includes long study in universities after finishing college.

7. Phil said: "Machines are taking over so much of our work that they will someday replace scientists." Phil's statement is wrong because

A. machines cannot build other machines.
B. men cannot let machines take over the world.
C. men cannot let machines run by themselves.
D. machines cannot think up new ideas.

8. When a scientist has a day off, he would probably not like to

A. go to his laboratory.
B. spend some time on his hobby.
C. go to a friend's party.
D. spend some time with his family.

9. The scientists of today can work on more complex problems than the scientists of the past mainly because they

A. work much harder than earlier scientists.
B. have more ideas than earlier scientists.
C. build on the work of earlier scientists.
D. are more clever than earlier scientists.

10. Most important scientific ideas are developed today as a result of

    A. a long study by a scientist working alone.
    B. work and thinking carried on by many scientists.
    C. team-work between scientists and the government.
    D. a group of scientific experts deciding what to study.

11. In observing and experimenting, a scientist today almost always needs

    A. computers and other large machines.
    B. many trained people to help him.
    C. microscopes, telescopes, and test tubes.
    D. different kinds of special equipment.

12. Scientists are often said to be very hardworking and quite devoted to their jobs. This is true of

    A. successful people in almost all kinds of work.
    B. scientists, but not people in other kinds of work.
    C. most people in important work, but not scientists
    D. more scientists than people in other kinds of work.

13. When many new facts are discovered which do not fit a scientific theory, scientists will usually ask themselves:

    A. Shall we throw out the theory since the facts do not fit it?
    B. Can we change the facts a little so that they will fit the theory?
    C. Shall we keep the theory as it is, since these new facts don't help it?
    D. Can we change the theory a little so that these new facts will fit in?

14. Scientists are most likely to make important discoveries by

    A. making many observations.
    B. trying out ideas.
    C. reading about experiments.
    D. asking many questions.

15. Before a scientist announces a new theory to the public, he will most likely talk his ideas over with

    A. other scientists in his special field.
    B. newspaper reporters who write about science.
    C. a group of experts on scientific theories.
    D. government leaders interested in his theory.

16. Which of the following sentences about science is best?

    A. Modern science is too advanced to use past discoveries.
    B. Modern science develops modern products.
    C. Modern science depends on useful inventions.
    D. Modern science is based on the science of the past.

17. The main reason that American scientists are asking for much money for research is that

    A. as much money should be spent on scientific work as on other things.
    B. many scientific projects take a long time and need much equipment.
    C. the cost of scientific materials and equipment has gone up since 1945.
    D. new scientific laboratories must be large and are expensive to build.

18. When a scientist reads a report of a new scientific discovery, he will probably

(continued)

18. (continued)

A. not fully believe the report until he has checked the work himself.
B. believe the report without asking too many questions about it.
C. not fully believe the report until he has obtained more information.
D. believe the report because it describes the work of scientists.

19. Scientists make many measurements in their work. Of the following, a scientist would be most likely to measure

A. how many germs a certain toothpaste kills.
B. the biggest load that a bridge can carry safely.
C. how far a bird flies south for the winter.
D. the distance a car runs on one gallon of gasoline.

20. Mary likes science. At first, she did not like to write down all the details of her experiments. If Mary becomes a scientist, however, this training will help her to

A. be patient in doing her experiments.
B. make better reports about her experiments.
C. develop theories from her experiments.
D. think up new experiments to perform.

21. Different groups of people help mankind in different ways. What is the special way in which scientists help mankind?

A. Scientists make better things for better living.
B. Scientists show us how to be more healthy.
C. Scientists give us knowledge about nature.
D. Scientists offer skilled service and advice.

22. When a scientist completes a new scientific theory, we know that he has

    A. created one of the laws of nature.
    B. helped bring mankind closer to the truth.
    C. discovered new ways of experimenting.
    D. developed new ideas and understandings.

23. The newest microscopes make it possible for scientists to study very small objects and also to

    A. explore many new problems.
    B. look for the meaning of life.
    C. observe atoms in motion.
    D. see that germs cause disease.

24. A scientific law describes

    A. rules which scientists must obey.
    B. rules which connect events in nature.
    C. rules for doing good experiments.
    D. good guesses about how things happen.

25. Which one of the following sentences best describes science?

    A. Science is experimenting.
    B. Science is planning and thinking.
    C. Science is thinking and doing.
    D. Science is observing and measuring.

26. A scientist is open-minded about his work if he

    A. discusses most of his ideas with others.
    B. considers ideas which go against his own.
    C. thinks up many new ideas for experiments.
    D. agrees with the ideas of other scientists.

27. Scientists study plants mainly to

A. help farmers to produce more food.
B. discover how to make new medicines.
C. understand how they live and grow.
D. find out where they will grow best.

28. Which of the following is the main need of science?

A. People with new ideas.
B. More money and equipment.
C. Well-trained workers.
D. Better working conditions.

29. When we say that a scientist has formed a hypothesis about an experiment, we mean that he has

A. indicated which measurements were made.
B. designed equipment needed for the experiment.
C. described how the experiment turned out.
D. made a careful guess about what will happen.

30. Which of the following is the best list of what scientists study?

A. Atoms, molecules, and stars.
B. Matter, energy, and living things.
C. Living things, disease, and growth.
D. Rockets, satellites, and space travel.

31. Bill always gets good grades in school, likes to build model airplanes, and plays jokes on his classmates. Frank gets high grades in arithmetic, likes to read books, and plays baseball. Janet is serious and smart, and likes to dance. Who do you think could become a scientist?

A. Bill
B. Frank
C. Janet
D. Any one of the three

32. Should a person who makes plans to build new types of airplanes be called a scientist?

A. Yes, because he uses scientific methods in his plans.
B. No, because he is planning to build a useful machine.
C. Yes, because he does experiments to check his plans.
D. No, because he is planning to try out some new ideas.

33. A scientific theory should

A. provide the final answer to scientific questions.
B. supply directions for making useful things.
C. tie together and explain many natural events.
D. suggest good rules for carrying out experiments.

34. When a scientist makes a new discovery, he usually makes a report of it. He does this because he

A. hopes other scientists will help him to finish his work.
B. wants other scientists to learn about his work and check it.
C. hopes to help his fellow man by announcing his discovery.
D. wants to know if others have done the same work and have reported it.

35. Experiments are used in science to

A. solve the problems of man.
B. find out the truth about nature.
C. try out the ideas of scientists.
D. prove that the universe is orderly.

36. "Most scientists are smart. They learn more easily than most people and can do harder things with their minds." Is this statement correct?

A. Yes, but scientists are generally no smarter than doctors or lawyers.
B. Yes, but only because scientists are born with scientific skills.
C. Yes, but only because scientists have received special training.
D. No. Scientists are about as smart as most people but no smarter.

This is the end of the questions in this booklet. If you finish before time is called, please go back and check your answers.

THE INQUIRY EFFICIENCY TEST

(An adaptation of the TAB Science Test)

A brief description of the films used in the test is given below. This description is placed here for the convenience of the reader and does not form part of the test papers received by the students.

Boiling Water by Cooling

A pyrex flask approximately half full of water is heated until the water boils. The flask is removed from the heat source and stoppered as soon as the water ceases boiling. An ice pack is placed on the outside of the flask. The water inside the flask boils once again.

The Bimetallic Strip

A bimetallic strip is heated and it bends. It is then immersed in cold water and it straightens out.

Page 1

One day in class Mr. Jones asked Randy to think of a number. What number was it?

A 9

B 47

C 52

D 150

Page 2

A1	Was the number greater than 100	NO
B1	Was the number an even number?	YES
C1	Was the number less than 9?	NO
D1	Was the number greater than 4?	YES
E1	Was the number an odd number?	NO
A2	Was the number less than 100	YES
B2	Was the number less than 52?	NO
C2	Was the number divisible by 3?	NO
D2	Was the number greater than 10?	YES
E2	Was the number 150?	NO

Page 3

The number that Randy was thinking of was:

A1	47	NO
B1	9	NO
C1	150	NO
D1	52	YES

Your answer sheet looks like this:

1. A B $\overset{(1)}{C}$ D E 36. A B $\overset{(2)}{C}$ D E 71. A B $\overset{(3)}{C}$ D E 106. A B $\overset{(4)}{C}$ D E
2. A B C D E 37. A B C D E 72. A B C D E 107. A B C D E

WE WILL USE ONLY NUMBERS 1 TO 35 AND THEY REFER TO THE COLUMNS RIGHT ACROSS THE PAGE. THE COLUMN HEADED BY NUMBER 36 IS THE SECOND COLUMN, THE COLUMN HEADED BY 71 IS THE THIRD COLUMN. IN THE ILLUSTRATION ABOVE, ONE PUPIL GUESSED THAT RANDY WAS THINKING OF 47 AND THEN AS HIS FIRST CLUE HE ERASED THE ANSWER FOR D2.

NUMBER 1 FOR THIS SECTION

PROBLEM I

WHY DOES THE WATER BOIL THE SECOND TIME?

Mark your selection on number 1.

A. Suction made the water boil the second time.

B. When a hot substance was placed over the bottle with the cork in it, it heated the water in the bottle and made it boil the second time.

C. When the substance was poured over the bottle, all of the steam that was in the bottle went out of the top of the bottle. Then, because the substance that was placed over the bottle was hot, the water inside the bottle boiled the second time.

D. When the cold substance was placed over the bottle with the cork in it, it mixed with the hot water and steam in the bottle made it boil the second time.

E. When the cold substance was placed over the bottle with the cork in it, the steam in the bottle cooled and turned back to liquid water. Then, because there was less pushing on the water and the water was still hot, the water boiled the second time.

After you have made your selection, turn the page and gather clues about the problem.

NUMBERS 2 TO 11 FOR THIS SECTION

Here are some clues you might want to gather to help you solve the problem. If you want to know the answers to any of the clue questions, erase the black mark covering the YES or NO answer opposite the question. Record your first choice as number 2, your second as number 3, and so on. Numbers 2 - 11 on your answer sheet are to be used for this section. ERASE ANSWERS ONLY FOR THOSE ANSWERS WHICH YOU THINK WILL HELP YOU. RECORD EACH CHOICE IN THE ORDER IN WHICH THEY WERE SELECTED.

A1. Was the bottle and everything in it cooler after the substance was placed over the bottle? YES

B1. When the water was boiled the first time, did the steam that came from the water push most of the air out of the bottle? YES

C1. Was the air in the bottle above the water pushing on the water more when the water boiled the first time than it was when it boiled the second time? YES

D1. Was the water in the bottle still hot when it boiled the second time? YES

E1. When the cork was put in the bottle, was the bottle filled mostly with steam? YES

A2. If no cork had been put in the bottle, would the water have boiled the second time? NO

B2. Was the substance that was placed on the bottle after the cork was put in the top, made up of ice wrapped in cloth? YES

C2. Did the substance that was placed over the bottle cause most of the steam in the bottle to cool and turn back to liquid water? YES

(cont'd)

D2.	Will water boil at a lower temperature if the air above it does not push down on it as hard?	YES
E2.	If the air in the bottle did not push down on the water as hard, would the water boil at the same temperature?	NO

NUMBERS 12 TO 15 FOR THIS SECTION

Read the selection over carefully. If you are fairly sure you know the right answer, erase the black mark covering the answer. If you get a "NO" response, then you may go back and gather more information by erasing more tabs from the opposite page. Keep going until you get a "YES" response on this page. Numbers 12 - 15 are to be used for this section.

A1. When the substance was placed over the bottle, all of the steam that was in the bottle went out of the top of the bottle. Then because the substance that was placed on the bottle was hot, the water inside the bottle boiled the second time. NO

B1. When the cold substance was placed on the bottle with the cork in it, the steam in the bottle cooled and turned back to liquid water. Then because there was less pushing on the water and the water was still hot, the water boiled the second time. YES

C1. When the hot substance was placed on the bottle with the cork in it, it heated the water in the bottle and made it boil the second time. NO

D1. Suction made the water boil the second time. NO

E1. When the cold substance was placed on the bottle with the cork in it, it mixed with the hot water and steam in the bottle and made it boil the second time. NO

STOP

Do Not Turn
the Page Until
You are Told To
Do So.

NUMBER 16 FOR THIS SECTION

PROBLEM II

WHY DOES THE BLADE BEND AND THEN STRAIGHTEN OUT?

Select one answer only for this page. Indicate your choice on number 16 on answer sheet.

A. The knife looks like this:

It is an ordinary table knife like a butter knife. It melts when it is put in the fire and bends. When the knife is put in the liquid, it cools and goes back to its normal shape.

B. The knife looks like this:

When the knife is heated, the top metal melts and causes the knife to bend. When the knife is put in the liquid, it cools and goes back to its normal shape.

C. The knife looks like this:

It is an ordinary knife like a butter knife. When it is put in the fire, the knife expands and bends. When the knife is put in the liquid it cools and goes back to its normal shape.

D. The knife looks like this:

When the knife is heated in the flame, it bends toward the ground because the bottom metal gets smaller and the top metal gets larger. When the knife is put in the liquid, it cools and goes back to its normal shape.

E. The knife looks like this:

When the knife is heated, the metals expand the same amount and the knife bends. When the knife is put in the liquid, it cools and goes back to its normal shape.

(continued)

A2. The knife looks like this:

When the knife is heated, the top metal expands faster than the bottom metal. The bottom metal then pulls the top metal into a curve and the knife bends. When the knife is put in the liquid, it cools and goes back to its normal shape.

After you have marked the proper space for your selection on number 16 on the answer sheet, turn the page and gather clues about the problem.

NUMBERS 17 TO 30 RESERVED FOR THIS SECTION

Here are some clues you might want to gather to help determine the correct solution. If you want to know the answer to any one of the clue questions, erase the black material as before and record your choice on the answer sheet. First choice number 17, second choice number 18 and so on. Select only those which you need. When you are sure you know the answer proceed to the next page.

A1. Was the knife made of brass and steel? YES

B1. Look at this knife:

Suppose the knife was heated and the top metal expanded faster than the bottom metal, would the bottom metal pull the top metal so that the knife bends? YES

C1. Was the knife longer when it was curved than when it was straight? YES

D1. After the knife was bent, if it would have been allowed to cool without putting it into the liquid, would it have straightened out? YES

E1. Was one side of the knife made of one kind of metal and the other of another kind of metal? YES

A2. Was the knife made of more than one kind of metal? YES

B2. Was the liquid in the tank such that it cooled the knife? YES

C2. If the knife were made of tin and gold, would the knife bend when it was heated? YES

D2. Do all metals expand the same amount if they are heated with the same amount of heat? NO

(continued)

E2.	Was the liquid in the tank water?	YES
A3.	Does one part of the knife expand faster than the other part of the knife?	YES
B3.	Does the knife expand when it is put in the flame?	YES
C3.	Suppose the knife was made of two different metals: would the knife bend if the two metals were attached to each other?	NO
D3.	Suppose the knife bent like this: does this mean that the metal on the top expands faster than the bottom metal if the knife is heated?	YES

NUMBERS 31 TO 35 RESERVED FOR THIS SECTION

WHY DOES THE BLADE BEND AND THEN STRAIGHTEN OUT?

Erase the black material covering the YES or NO beside the answer that you think answers the question most completely. Record your choice on the answer sheet. Your first choice will be number 31, your second number 32 and so on. If your first choice is a "NO" response then you may go back to the previous page and gather more clues. Keep going until you get a "YES" answer on this page.

A1. The knife looks like this:

When the knife is heated in the flame, it bends toward the ground because the bottom metal gets smaller and the top metal gets larger. When the knife is put in the liquid, it cools and goes back to its normal shape. NO

B1. The knife looks like this:

When the metal is heated, the top metal expands faster than the bottom metal. The bottom metal then pulls the top metal into a curve and the knife bends. When the metal is put in the liquid it cools and goes back to its normal shape. YES

C1. The knife looks like this:

It is an ordinary knife, like a butter knife. When it is put in the fire, the knife expands and bends. When the knife is put in the liquid, it cools and goes back to its normal shape. NO

D1. The knife looks like this:

When the knife is heated, the metals expand the same amount and the knife bends. When the knife is put in the liquid, it cools and goes back to its original shape. NO

(continued)

E1. The knife looks like this:

When the knife is heated, the top metal melts and causes the knife to bend. When the knife is put in the liquid, it cools and goes back to its normal shape. NO

After you have found the "YES" answer on this page, STOP. CLOSE THE TEST BOOKLET. DO NOT GO BACK TO OTHER PARTS OF THE TEST.

SCIENCE REASONING TEST (SRT)

Section I

This section of the test consists of matching the underlined parts of a paragraph with the items in a KEY as illustrated in the following example:

KEY

A. A boy's name
B. A girl's name
C. The name of a country
D. A kind of bird
E. A kind of tree

He walked down the road until he met (1) John. They continued on their way until they reached an open field. There they saw a (2) robin perched on the branch of a (3) poplar tree.

For this example you would mark your answer sheet as shown below:

1. A ■ 1 B = 2 C = 3 D = 4 E = 5

2. A = 1 B = 2 C = 3 D ■ 4 E = 5

3. A = 1 B = 2 C = 3 D = 4 E ■ 5

The following paragraphs are about scientific investigations that took place a long time ago. See if you can mark your answer sheet so as to match the underlined parts with the items in the KEY below.

KEY

A. A problem or question
(The problem may not be stated as a question)
B. Hypothesis or a guess about the solution to the problem
C. Observations which were made before the experiment took place
D. Results of experimenting
E. An explanation or conclusion

Some people had noticed that when they were combing their hair, (1) the hair was attracted to the comb and this comb would then attract little pieces of paper. A scientist, Charles Dufay, gave a glass rod an electric charge by rubbing the rod with silk. He charged a gold leaf with electricity by touching it to the glass rod. To his surprise he noticed that (2) the gold leaf was pushed away or repelled by the glass rod. (3) "I had expected the opposite effect," he said, "but I now believe (4) there are two kinds of electricity."

"Having considered that (5) flies swarmed around decaying flesh, I began to believe that (6) All worms found in flesh came from the eggs of the flies and not from the decaying of the meat. I placed different kinds of flesh in glass jars and sealed these jars so that nothing could enter these jars. I then filled an equal number of jars the same way but left these open. The flesh decayed in all of the flasks but (7) worms were found only in the open flasks. I repeated experiments such as these a large number of times. I am now convinced that (8) the flesh of dead animals cannot produce worms unless the eggs of some insect had somehow been placed on it.

KEY

A. A problem or question (the problem will not usually be stated as a question.)
B. Hypothesis (a guess about the answer to the problem)
C. Observations which were made before the experiment took place
D. Results of experimenting
E. An explanation or conclusion

He began his experiments thinking that it was likely that (9) heating a substance had no effect on its weight. He watched men working in a shop and noticed that (10) when holes were being drilled in metal, the metal became quite hot. He obtained a piece of metal and weighed it carefully. He then drilled holes in the metal and again weighed the metal plus the metal chips. (11) No change in weight was found. He repeated these experiments using different materials until he was convinced that (12) heat was a form of motion and nothing else.

Section II

This section of the test consists of matching an experiment with the items in a KEY. For each problem below one experiment is the best one. The others are unsatisfactory for some reason or another. Some of these experiments may be unsatisfactory for the same reason as others. YOU ARE TO JUDGE THE EXPERIMENTS ACCORDING TO THE FOLLOWING KEY:

KEY

A. This is the best experiment.
B. This experiment is not satisfactory because it does not have a control or something to compare it with.
C. This experiment does not have a very good control or comparison.
D. This experiment is not testing the hypothesis or guess.
E. This experiment is unsatisfactory because of other reasons not mentioned above.

PROBLEM What are some of the things needed for the sprouting of seeds?

HYPOTHESIS (GUESS) Oxygen is one of the requirements for the sprouting of seeds.

13. Plant one seed in a container of moist soil in which oxygen is present and plant another seed in a container in which oxygen is not available.

14. Plant 100 seeds in a container of moist soil. Remove all of the oxygen.

15. Plant 100 seeds each in two separate containers of moist soil. Place one container in the sunlight and the other in darkness. Keep all other conditions the same.

16. Plant 100 seeds each in two separate containers of moist soil. One container contains air from which only the oxygen has been removed while the other contains ordinary air.

17. Plant 100 seeds each in two separate containers of moist soil. Remove all of the air from one container while the other container contains ordinary air.

KEY

A. This is the best experiment.
B. This experiment is not satisfactory because it does not have a control or something to compare it with.
C. This experiment does not have a very good control or comparison.
D. This experiment is not testing the hypothesis or guess.
E. This experiment is not satisfactory because of other reasons not mentioned above.

PROBLEM What is needed for the production of starch in the leaves of green plants?

HYPOTHESIS Light is needed for the production of starch in the leaves of plants.

18. Place one plant in sunlight and a similar plant in darkness. Keep all other conditions the same. After a few days, test the leaves for starch.

19. Place several plants in sunlight. Cover one of the plants with a light-proof cover so as to keep out all of the light. Keep all other conditions the same. After a few days test the leaves for starch.

20. Place several plants in sunlight. Cover half of the plants with a light-proof cover so as to keep out all of the light. After a few days test the leaves for starch.

21. Select several different plants which have been in sunlight for a few days. Test the leaves for starch.

22. Select several plants and place them in sunlight. Place the same number of the same kind of plants under the light from ordinary light bulbs. Leave both sets of plants for a few days and then test the leaves for starch.

Section III

In this section you are asked to look at some facts and decide whether these facts support a guess or hypothesis about the solution to the problem. Look at the example below:

PROBLEM Who broke the window?

HYPOTHESIS George broke the window.

Facts:

1. George was not near the window shortly after it was broken. (This would be indirect evidence against the hypothesis.)
2. George said that some day he would break the window. (This would be indirect evidence to help prove the hypothesis.)
3. George threw a brick at the window. (This would be direct evidence to help prove the hypothesis.)

NOW TURN TO THE NEXT PAGE AND MATCH THE ITEMS IN THE KEY WITH THE FACTS IN NUMBERS 23 TO 35.

In this section a problem is given followed by two hypotheses or guesses. SEE IF YOU CAN JUDGE HOW WELL THESE FACTS HELP TO PROVE OR NOT PROVE THE GUESS OR HYPOTHESIS.

KEY

A. This fact offers direct evidence to help prove the hypothesis.
B. This fact offers indirect evidence to help prove the hypothesis.
C. This fact has nothing to do with the guess or hypothesis.
D. This fact offers indirect evidence against the hypothesis.
E. This fact offers direct evidence against the hypothesis. (It helps to prove that the hypothesis is not true.)

PROBLEM What causes Disease A?

HYPOTHESIS I In man the disease is spread from one person to another.

Facts: (for Hypothesis I)

23. People, looking after patients suffering from this disease, did not catch the disease.

24. People, living in areas where the disease is common, often caught the disease.

25. The disease is not common in Canada.

26. Five prisoners were kept in one room. One of them was ill with disease A. One of the well prisoners caught the disease.

27. Monkeys suffering from this disease do not give it to other monkeys.

HYPOTHESIS II The disease is spread by a species of mosquito common to tropical areas.

Facts: (for Hypothesis II)

28. The disease is rare in cool or temperate climates.

29. The disease is common in many tropical areas.

30. People suffering from the disease often have a very high fever.

31. The disease becomes less common after swamps are drained.

32. The parasite of the disease was found living on these tropical mosquitoes.

33. Two people stung by these mosquitoes did not catch the disease.

34. Some men taking part in an expedition near swamps became ill with the disease.

35. Several mosquitoes of this species showed no evidence of the disease A parasite.

CHECK OVER YOUR PAPER IF YOU HAVE TIME. HAND IN YOUR ANSWER SHEET AND QUESTION PAPER.

SCIENCE REASONING TEST (SRT) KEY

1. C
2. D
3. B
4. E
5. C
6. B
7. D
8. E
9. B
10. C
11. D
12. E
13. E
14. B
15. D
16. A
17. C
18. E
19. E
20. A
21. B
22. D
23. D
24. B
25. C
26. A
27. D
28. B
29. B
30. C
31. B
32. A
33. D
34. B
35. E

THE PROCESSES OF SCIENCE TEST

This test consists of two short films of actual investigations. Each film is followed by a number of questions. Each question contains four possible reponses. You are to select the best or most complete answer from these responses.

EXAMPLE:

1. Dogs:

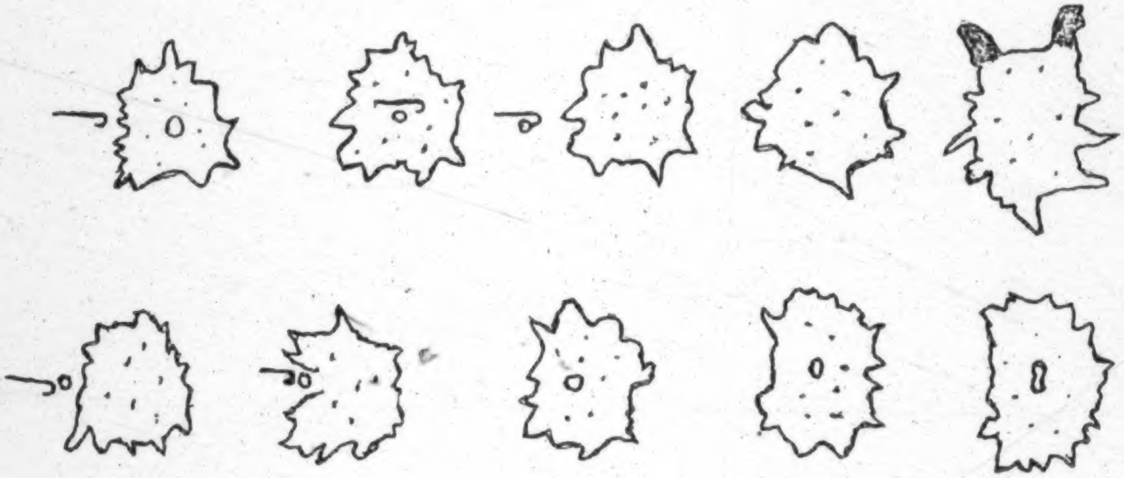
A. are animals
B. are animals with four legs
C. belong to the Cat Family
D. are warm-blooded animals with four legs

In this example, A, B, and D are correct answers. However D is the most complete answer so we would mark our answer sheet thus:

1. A. __ B. __ C. __ D. ■ E. __

NOTICE THAT THERE ARE ONLY FOUR POSSIBLE RESPONSES FOR EACH QUESTION. SO DO NOT PUT ANY MARKS FOR E ON YOUR ANSWER SHEET.

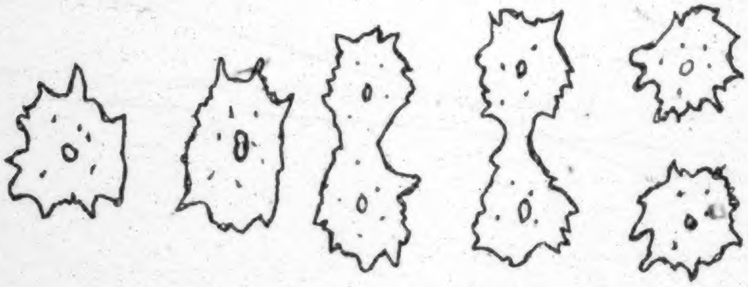
THE AMOEBA - WITH AND WITHOUT A NUCLEUS AS SHOWN IN THE FILM



The diagrams drawn above show some of the main parts of the investigation shown in the film. They are shown here to help you remember what you saw in the film.

THE AMOEBA - AS IT GROWS AND DEVELOPS UNDER NATURAL CONDITIONS

This single-celled animal grows larger, then divides into two new single-celled animals.



THE AMOEBA

1. What is the main problem? Select your answer from the list below.

   A. Can an amoeba without a mucleus divide into two parts?
   B. Does the nucleus have a special use?
   C. Do all amoeba have nuclei?
   D. Does the nucleus control how the single-celled animal grows?

2. To get at the main problem an investigation is carried out. The first part of the investigation attempts to find out:

   A. Whether an amoeba contains a nucleus.
   B. If a nucleus is needed for the amoeba to divide into two parts.
   C. If a nucleus is needed for the animal to live and grow in a normal way.
   D. If a nucleus is needed for the amoeba to remain alive.

3. In the first part of the investigation, the control was:

   A. The apparatus used in the experiment.
   B. The amoeba without a nucleus.
   C. The amoeba with a nucleus.
   D. The amount of movement possible for the amoeba.

4. The hook used to move the nucleus around:

   A. Was so small that it could be seen only with a microscope far more powerful than a school microscope.
   B. Was of approximately the size of a pin.
   C. Was so small that it had to be handled with a special apparatus.
   D. Was moved automatically as the nucleus moved.

5. If this first part of the investigation was carried out a large number of times, you would probably group your data under the headings:

   A. (1) Protoplasm (2) Nucleus (3) Amoeba
   B. (1) Amoeba with nuclei (2) Amoeba with nuclei removed
   C. (1) Number of amoeba that divided into two parts (2) Number of amoeba that survived
   D. (1) Amoeba with protoplasm (2) Amoeba without protoplasm

6. In the film it appeared that the nucleus of one amoeba was pushed into an amoeba from which the nucleus had been removed. This was done to:

   A. Decrease the effect of the nucleus in the experiment.
   B. See whether the nucleus of one cell could be transplanted into another.
   C. See whether an injured cell would remain alive.
   D. Make sure that the amount of injury would be about the same for the cell with a nucleus and the cell without a nucleus.

7. From this experiment:

   A. We have proved that the nucleus is needed for the cell to grow and divide in a normal way.
   B. We have evidence that the nucleus is necessary for the amoeba to remain alive.
   C. We have evidence to support the hypotheses that the nucleus has something to do with growth and cell division.
   D. We have shown that the nucleus helps to heal the injured cell.

8. Suppose that out of 100 amoeba which had their nuclei removed, two of them continued to grow and subdivide in a normal way, you would likely decide:

   A. Approximately 1% of the amoeba do not need nuclei for growth and subdivision.
   B. Approximately 4% of the amoeba do not need nuclei for growth and subdivision.

C. An imperfect job had been done and the nuclei had not been completely removed.
D. The two amoeba had probably captured new nuclei from the water.

9. From what you have seen you would expect the nucleus to contain:

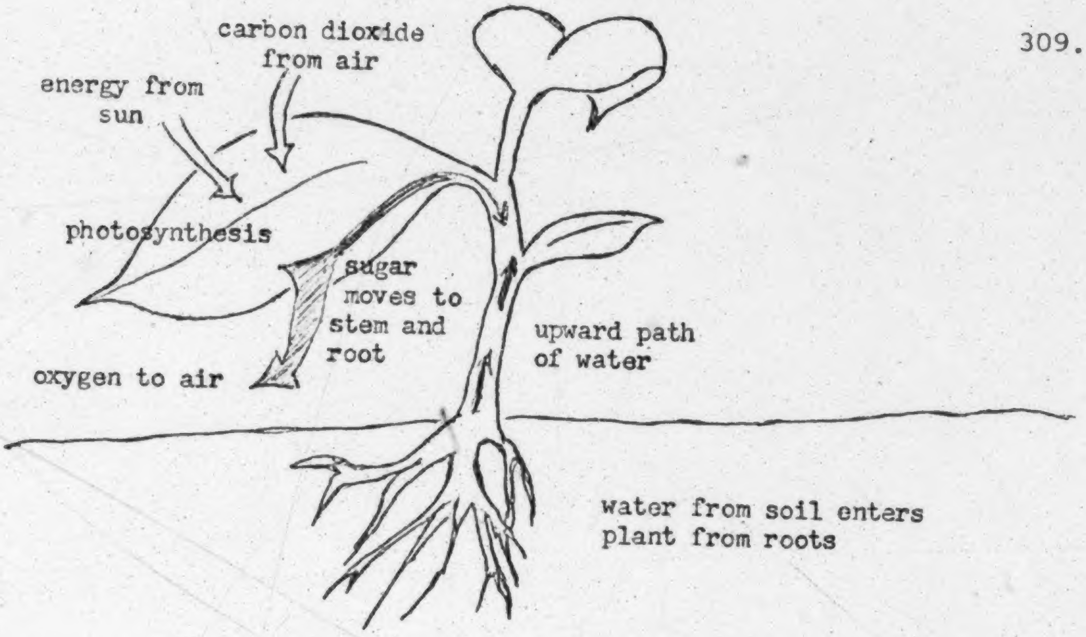
A. A substance that caused the amoeba to grow.
B. A substance that causes the amoeba to divide into two parts.
C. One or more substances which regulates how an amoeba grows.
D. A condensed part of the protoplasm but otherwise, the same as the rest of the cell.

10. What problem is suggested from the evidence you have seen?

A. Does injury stop growth of the amoeba?
B. How long will an amoeba survive without a nucleus?
C. Does an amoeba require its original nucleus in order to survive?
D. Can the nucleus of an amoeba be transplanted into another amoeba from which the nucleus had been removed?

11. Such operations as kidney transplants work well between close relatives (such as twin sisters). Using this information and the evidence from the film, choose the most likely statement.

A. All descendents of an amoeba are similar.
B. All descendents of an amoeba are similar unless affected by outside influences such as radiation.
C. The descendents of amoeba are quite different after each cell division.
D. There is no evidence of similarity or dissimilarity of the descendents of an amoeba.


carbon dioxide from air
energy from sun
photosynthesis
sugar moves to stem and root
upward path of water
oxygen to air
water from soil enters plant from roots

THIS DIAGRAM ILLUSTRATES THE PROCESSES GOING ON IN ALL GREEN PLANTS WHICH MANUFACTURE THEIR OWN FOOD. THE SUN'S ENERGY IS NECESSARY FOR THIS TO HAPPEN. IN THE FILM WHICH WE WILL LOOK AT, THE REACTION OF PLANTS TO SUNLIGHT IS ILLUSTRATED. REMEMBER THAT ALTHOUGH ONLY A FEW PLANTS ARE SHOWN IN THE FILM, MANY PLANTS ARE USED IN THE INVESTIGATION.

WATCH THE FILM CAREFULLY AND SEE IF YOU CAN RECOGNIZE THE MAIN PROBLEM, HYPOTHESES, ASSUMPTIONS, CONTROLS, USED IN THE INVESTIGATION, AND SO ON. A NUMBER OF QUESTIONS WILL BE GIVEN TO SEE HOW WELL YOU UNDERSTAND THIS INVESTIGATION. SELECT THE BEST ANSWER IN EACH QUESTION.

REACTION OF PLANTS TO LIGHT

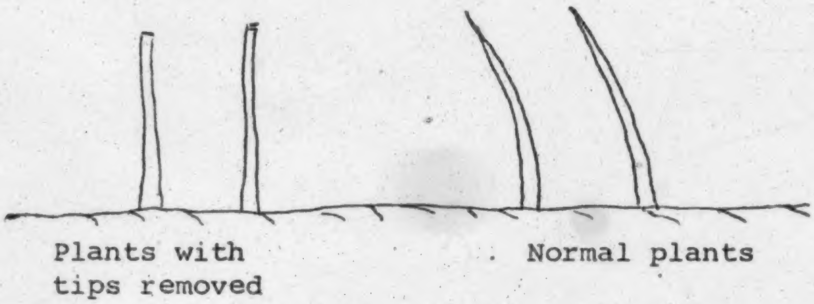
12. The first hypothesis to be tested was:

    A. The top of the seedling had something to do with the reaction.
    B. The root of the seedling had something to do with the reaction.
    C. Phototropism is caused by sunlight.
    D. The cells on the stem grow faster on the side away from the sun.

13. The seedlings were grown in the dark because:

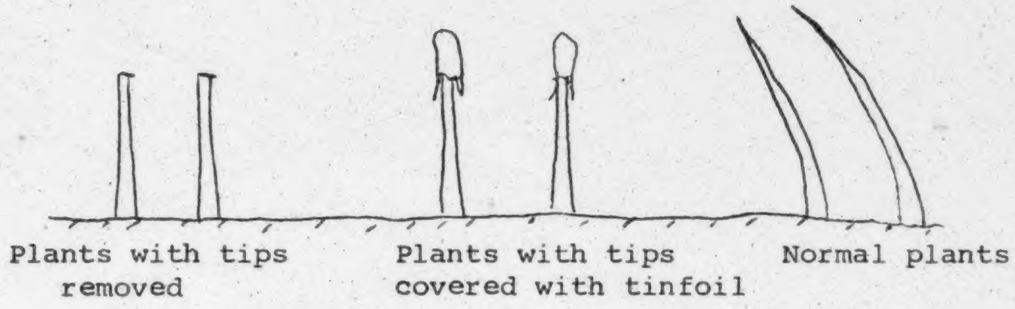
    A. They germinate better in the dark.
    B. The experimenter did not want the sunlight to affect the plant before the investigation started.
    C. The light causes the plants to bend.
    D. The plants grown in sunlight acts as a control.

14.



This part of the experiment shows that:
A. The tip of the plant causes the plant to react to light.
B. Plants with tips removed do not react to light.
C. Oat seedlings with tips removed do not react to light.
D. The cells on the side away from the sun are longer than the cells on the side facing the sun.

15.


Plants with tips removed
Plants with tips covered with tinfoil
Normal plants

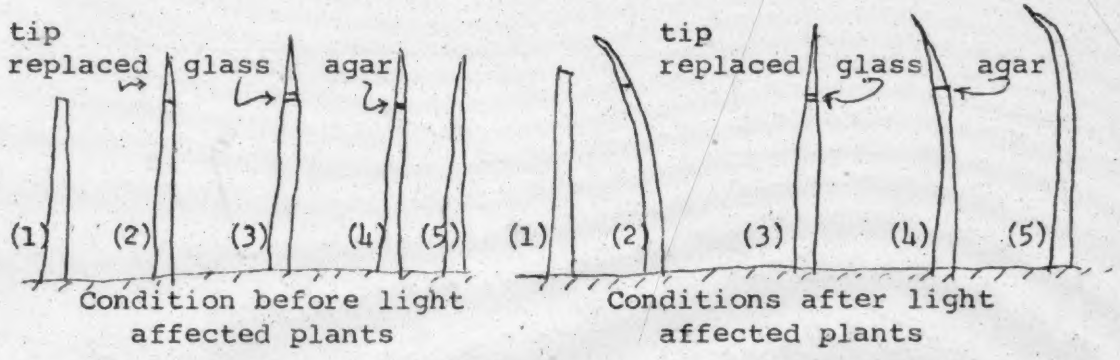
In this part of the investigation:

A. There is no evidence that the tip of the plants is involved.
B. The plants with foil caps on the tips act as a control for the movement of fluids from the tip of the plant.
C. The plants with the foil cap act as control for injury to plant and movement of fluids from the tip.
D. This part of the investigation had no controls.

16. The main problem in this investigation is:

A. Is light necessary for plant growth?
B. What causes plants to react to light?
C. Is sunlight necessary for plant growth?
D. What does the passage of fluids have to do with the plant's reaction to light?

Questions 17, 18 & 19 below refers to the part of the investigation illustrated by the following diagram.


tip replaced
glass
agar
(1)
(2)
(3)
(4)
(5)
Condition before light affected plants
tip replaced
glass
agar
(1)
(2)
(3)
(4)
(5)
Conditions after light affected plants

17. The sample used as control for the whole investigation was:

    A. (1)
    B. (4)
    C. (5)
    D. (2)

18. The samples in which fluids could not pass from the tip to the rest of the plant were:

    A. (1) only.
    B. (1), (2), (3) and (4) only.
    C. (2), (3) and (4) only.
    D. (1) & (3) only.

19. The samples in which injury was present and in which fluids could not pass from the tip to the rest of the plants were:

    A. (2) and (4)
    B. (2) and (3)
    C. (1) and (3)
    D. (2) and (5)

20. From what you have seen you would expect:

    A. The tip of the plant to contain a substance which reacts to light.
    B. The tip of the plant is the only part of the plant involved in the reaction to light.
    C. The tip of the plant and the movement of fluids to and from the tip of the plant are involved in the reaction of the plant to light.
    D. The movement of fluids to and from the tip is involved in the reaction of the plant to light.

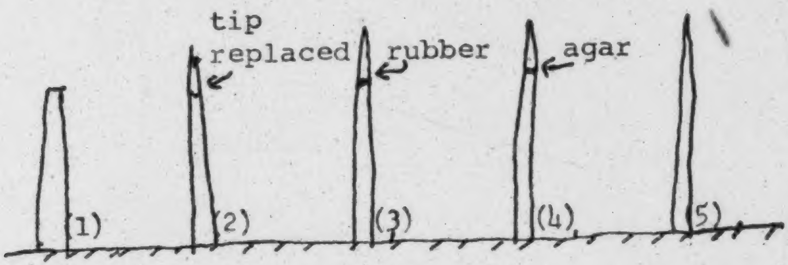
21. What problem is suggested from the evidence you now have?

    A. Do oat plants bend toward the sunlight?
    B. Are the cells of the plant the same size on the side nearest the sun as those on the side away from the sun?

21. (cont'd)

C. Can liquids pass through agar?
D. Do plants other than oats behave the same way?

22. I the diagram below in which wheat plants were used, edict which plants would bend toward the light?



A. (1), (5) and (3) only.
B. (3) and (4) only.
C. (2), (4) and (5) only.
D. (2) and (5) only.

WHEN YOU HAVE COMPLETED THIS SECTION CHECK OVER YOUR PAPER TO MAKE SURE YOU HAVE RECORDED ALL YOUR ANSWERS ON YOUR ANSWER SHEET, WRITTEN YOUR NAME AND CLASS (7-A, 7-B, 7-C OR 7-D etc.) IN THE PROPER SPACE. HAND IN YOUR QUESTION PAPER, ANSWER SHEET AND PENCIL.

PROCESS OF SCIENCE TEST (PST) KEY*

1. B
2. D
3. C
4. C
5. B
6. D
7. C
8. C
9. C
10. B
11. B
12. A
13. B
14. C
15. C
16. B
17. C
18. D
19. C
20. C
21. D
22. C

---

* The film loops (An Inquiry-The Importance of the Nucleus and Phototropism) which accompany the test are part of the BSCS series of Single Topic Inquiry Films. These film loops are available from commercial publishers which handle BSCS materials.

HOW I FEEL ABOUT MY SCHOOL

Developed by

J.K. Coster

I am conducting a survey of the opinion, beliefs and feelings of a large number of students like yourself. I would like to find out how you feel and what you think about your school. Your answers to the questions will provide us with information which will help us in improving our schools.

This is NOT a test. There are no right answers. YOUR OPINIONS OR ANSWERS ARE THE ONES IN WHICH I AM INTERESTED AND THE ONES THAT I WANT. I would like you to be very frank in making your selections. PLEASE DO NOT WRITE YOUR NAME ON THIS PAPER. WRITE ONLY YOUR IDENTIFICATION NUMBER ON THE SHEET. You are assured that no one in your school will see your papers. You may be certain, therefore, that the way in which you answer these questions definitely will NOT affect your standing in your school.

The questionnaire consists of thirty one questions about many parts of your school. Following each question is a list of five possible answers. Please read each question carefully. Then carefully read ALL five answers and select the ONE answer with which you agree most fully. (If none of the possible answers exactly express your opinion, you are asked to select the answer that most nearly expresses it.) After you have made your selection, record this selection on your answer sheet.

EXAMPLE: 1. On the average, is Alaska warmer or colder than Hawaii?

A. Alaska is much warmer than Hawaii.
B. Alaska is slightly warmer than Hawaii.
C. Alaska is neither warmer nor colder than Hawaii.
D. Alaska is slightly colder than Hawaii.
E. Alaska is much colder than Hawaii.

E is the best answer; therefore you would mark your answer sheet:

A — 1 B — 2 C — 3 D — 4 E 5

Please work rapidly but carefully. Answer *ALL* questions but select only *one* answer for each question. If you have any questions about the questionnaire you may ask the person or persons in charge. If you have no questions you may begin.

HOW I FEEL ABOUT MY SCHOOL

1. Do you believe that the work that you are taking in school will be useful to you after you leave school?

E. Yes! I am certain that my school work will help me a great deal.
D. I feel that my school work generally will be useful.
C. I feel that my school work will be of *some* use to me.
B. I feel that my school work will be of *little* use to me.
A. No! I am certain that my school work will be of No use to me.

2. Do you feel that your teachers treat you fairly?

E. Yes! My teachers *always* treat me fairly.
D. My teachers usually treat me fairly.
C. Sometimes my teachers treat me fairly; sometimes they don't.
B. My teachers usually do *NOT* treat me fairly.
A. No! My teachers *never* treat me fairly.

3. What is your opinion of your chances of getting the kind of job that you would like to have after you leave school (or college, if you plan to attend)?

A. I feel that I have a *very poor* chance of getting the kind of job I want.
B. I feel that I have a *poor* chance of getting the kind of job I want.

3. (continued)

C. I feel that my chances are about average.
D. I feel that I have a good chance of getting the kind of job I want.
E. I feel that I have an excellent chance of getting the kind of job I want.

4. How interesting, do you feel, is your school work to you?

E. All of my school work is interesting to me.
D. Most of my school work is interesting to me.
C. About half of my school work is interesting to me.
B. Little of my school work is interesting to me.
A. None of my school work is interesting to me.

5. What is your opinion of the working and studying conditions in your school?

A. Our working and studying conditions are very poor!
B. I feel that they are poor.
C. I feel that they are about average.
D. I feel that they are good.
E. Our working and studying conditions are excellent! They couldn't be better!

6. Do you feel that the other students in your school like you?

E. No! I feel that NONE of the other students in my school like me.
D. I feel that only a few of the other students like me.
C. Some of the other students like me; some don't.
B. I feel that most of the other students like me.
A. Yes! I feel that ALL of the other students in my school like me.

7. What is your general feeling of the school that you are now attending?

7. (continued)

E. There is no other school that I would like as well as this one.
D. Generally, I am satisfied with my school.
C. I feel that this is an average school.
B. Generally, I am NOT satisfied with this school.
A. I feel that this is the worst school that I could attend.

8. Do you believe that your parents are interested in your school work?

E. Yes! I am sure that my parents are highly interested in my school work.
D. On the whole, I feel that they are interested in my school work.
C. Sometimes I feel that they are interested; sometimes I feel that they aren't.
B. On the whole, I feel that they are NOT interested in my school work.
A. No! I am certain that my parents are NOT interested in my school work.

9. What is your opinion of the number of activities - such as clubs, dances, parties and sports - in your school?

A. I am greatly dissatisfied with the small number of activities.
B. On the whole, I feel that we DON'T have enough activities in this school.
C. We have a fair number of activities, but we should have more.
D. On the whole, I feel that we have enough activities in our school.
E. I am very well satisfied with the number of activities in our school.

10. What, in general, is your opinion of the teachers in your school?

    E. No other school has a better group of teachers than ours.
    D. I believe that we have a good group of teachers in our school.
    C. I feel that we have an average group of teachers in this school.
    B. I believe that we have a poor group of teachers in this school.
    A. No other school has a poorer group of teachers than this one.

11. What is your opinion of the group of subjects (or courses) that is offered in your school?

    A. No school student should have to study these subjects.
    B. I feel that this school offers a poor group of subjects,
    C. This school offers a fair group of subjects.
    D. I feel that this school offers a good group of subjects.
    E. I couldn't ask for a better group of subjects.

12. Do you feel that your school work is the kind of work that you like to do?

    E. Yes! All of my school work is the kind of work that I like to do.
    D. Most of my school work is the kind of work that I like to do.
    C. Some of my school work is the kind that I like to do; some is not.
    B. Little of my school work is the kind of work that I like to do.
    A. No! None of my school work is the kind of work that I like to do.

13. Do you believe that your school teachers are personally interested in you?

A. No! I feel that my school teachers are NOT interested in me.
B. Generally, I DON'T believe that they are interested in me as a person.
C. Sometimes I feel that my teachers are personally interested in me.
D. Generally, I believe that they are interested in me as a person.
E. Yes! I definitely feel that my teachers are interested in me as a person.

14. In your opinion, how do people in your community feel about your school?

A. The people are greatly dissatisfied with this school.
B. Generally, they are dissatisfied with this school.
C. About half of the people in this community are satisfied with the school.
D. Generally, they are satisfied with our school.
E. The people are very well satisfied with our school.

15. How do you feel about the way in which your school subjects are taught?

E. I feel that all of my subjects are taught in an excellent manner.
D. On the whole, I am satisfied with the way in which they are taught.
C. Some subjects are taught in a satisfactory manner; others are not.
B. On the whole, I am dissatisfied with the way in which they are taught.
A. I am highly dissatisfied with the way in which my subjects are taught.

16. What is your opinion of the help and assistance that your school teachers give you with your school work?

    A. My school teachers never help me.
    B. I feel that my school teachers give me very little help.
    C. Sometimes my teachers give me the help that I need.
    D. My teachers usually help me when I need it.
    E. I feel that my teachers are always very generous with their help.

17. How do you feel about going to adults in your school - such as teachers, principal, or counsellors - to get help and advice regarding your personal problems such as how to improve your appearance, how to act on a date, what subjects to take in school, or how to get along with other people?

    E. I feel that I definitely would want to get help with my problems.
    D. I frequently would want to get help from an adult in my school.
    C. I would want to get help with some, but not all, personal problems.
    B. I rarely would want to get help from an adult in my school.
    A. I am sure that I would NOT go to anyone in this school for help.

18. What is your opinion of the marking and grading system in your school; that is, how do you feel about the way in which marks or grades are given?

    A. I feel that the system in this school is very unsatisfactory.
    B. Generally, I feel that this school has a poor system.
    C. The system in this school may be all right, but I don't like it.
    D. Generally, I feel that our school has a good system.
    E. I feel that our school has an excellent system.

19. How do you feel about the way in which your school is organized?

    E. The organization is excellent; our school runs very smoothly.
    D. On the whole, I feel that our school is well-organized.
    C. The organization of our school is just "so - so".
    B. On the whole, I feel that this school is poorly organized.
    A. The organization is very poor; no one knows what is going on.

20. What is your opinion of the school spirit in your school?

    A. There is absolutely NO school spirit in this school.
    B. The school spirit is low.
    C. The school spirit is about average when compared with other schools.
    D. Our school spirit is high.
    E. Our school spirit is excellent! It couldn't be higher.

21. How do you feel about the social life you are having while attending school?

    A. I am greatly disappointed with my social life.
    B. Generally, I am dissatisfied with my social life.
    C. Sometimes I am satisfied with my social life; sometimes I am not.
    D. Generally, I am satisfied with my social life.
    E. I am very well satisfied with my social life.

22. Do you feel that going to school will help you in enjoying life more and in getting more satisfaction from living?

    A. No! I am certain that going to school will NOT help me.

22. (continued)

B. On the whole, I DON'T believe that going to school will help much.
C. Going to school may be of some help to me in enjoying life.
D. On the whole, I feel that going to school will be helpful.
E. Yes! I am sure that going to school will help me in enjoying life.

23. How hard do you feel that you are working (or studying) on your school work?

A. I never work in school.
B. Usually I DON'T work very hard in school.
C. Sometimes I work hard; sometimes I don't.
D. Usually, I work hard in school.
E. I always work as hard as I can on my school work.

24. What is your opinion of the other boys and girls in Your school?

E. They are the best group of boys and girls in the world.
D. I feel that we have a good group of boys and girls in our school.
C. Some of the other students are all right; some are not.
B. I feel that this school has a poor group of boys and girls.
A. They are the worst group of boys and girls in the world.

25. How well, in your opinion, do your school teachers "know" and understand the subjects that they teach?

A. My teachers definitely do NOT know and understand their subjects.
B. They don't "know" and understand their subjects as well as they should.

25. (continued)

C. Some of my teachers "know" and understand their subjects; some don't.
D. Generally, I feel that they "know" and understand their subjects.
E. My teachers definitely do "know" and understand their subjects.

26. What is your opinion of the equipment and facilities - such as laboratory and shop equipment, books and desks - in your school?

E. No other school has better equipment and facilities
D. Our equipment and facilities are better than those in most schools.
C. The equipment and facilities in our school are about average.
B. Our equipment and facilities are poorer than those in most schools.
A. No other school has poorer equipment and facilities than ours.

27. In general how do you feel that the other people in your school treat you?

E. I feel that I couldn't be treated better!
D. Generally I feel that I am treated in a satisfactory manner.
C. Sometimes I am treated all right; sometimes I feel that I am not.
B. Generally I feel that I am NOT treated very well.
A. I feel that I couldn't be treated worse!

28. In your opinion, how good a job does your school do in educating the students who come to it?

E. Very good or outstanding.
D. Good
C. Average
B. Poor
A. Very poor

29. How good a job do your teachers do in teaching you?

E. Very good or outstanding
D. Good
C. Average
B. Poor
A. Very poor

30. How well do you like your teachers, as persons, who teach you?

E. I like my teachers very much.
D. I like my teachers.
C. I neither like or dislike my teachers.
B. I don't like my teachers.
A. I dislike my teachers very much.

31. How good a job does your principal do as a principal?

E. Very good or outstanding
D. Good
C. Average
B. Poor
A. Very poor

The Observational Instrument

THE OBSERVATIONAL INSTRUMENT

Part 1: The Processes

Time (2 min. interval) Specific Behaviors Major Process

(a) Speculating about a phenomenon (b) Identifying variables (c) Noting and making assumptions (d) Delimiting the problem	Identifying a Problem	1
(a) Recalling knowledge and experience (b) Doing literature research (c) Consulting people	Background Information	2
Predicting		3
Hypothesizing		4
(a) Defining independent and control variables (b) Defining procedure and sequencing the steps (c) Identifying equipment, materials and technique (d) Indicating safety precautions (e) Devising the method for recording data	Design	5
(a) Collect, construct, and set up apparatus (b) Performing the experiment (c) Identifying limitations of design (d) Repeating the experiment (e) Recording data	Procedure	6

THE OBSERVATIONAL INSTRUMENT

Part 1: (continued).

(a) Obtaining qualitative data (b) Obtaining quantitative data (c) Gathering specimens (d) Obtaining graphical data (e) Serendipity (f) Noting precision and accuracy of data (g) Judging reliability and validity	Observing and Observations	7
(a) Ordering to identify regularities (b) Classifying (c) Comparing	Organizing Data	8
(a) Drawing graphs, charts, maps (b) Interpolating and extrapolating	Rep. Data Graphically	9
(a) Computing (b) Using statistics (c) Determining uncertainty in results	Treating Data Mathematically	10
(a) Suggesting an explanation for data (b) Deriving an inference or generalization (c) Assessing validity	Interpreting	11
(a) Verbal operational definitions (b) Mathematical operational definitions	Operational Definitions	12

THE OBSERVATIONAL INSTRUMENT

Part 1: (continued)

Mathematical relationship		13
Incorporating into theory		14
(a) Increase level of confidence (b) Test range of application	Seeking Further Evidence	15
(a) Effect of a new variable (b) Unexpected observations (c) Incompleteness in theory	New Problems	16
Applying the discovered knowledge		17

THE OBSERVATIONAL INSTRUMENT (continued)

Part II: The Characteristics of the Science Teacher

Time Interval

Gives direction Introduces Lectures Summarizes Explains	Teacher Talks	TEACHING TECHNIQUES
Recitation Request and answer questions Discussion	Teacher Student Talk	
Uses A-V aids Demonstration Helps individual student	Teacher Does	
One student - class Students work indv. Laboratory work	Students Do	
Review Evaluation	Purpose	
Recall facts See relationships Make observations Hypothesize Test hypothesis	Teacher's Questions	
Concrete Abstract	Method	CHARACTERISTICS
Practical Theoretical	Subject Matter	
Directed Non-directed	Pupil Activity	

INSTRUCTIONS FOR THE USE OF THE OBSERVATIONAL INSTRUMENT

Interpretations

Part 1: The Processes

Interpret these behaviors as outlined by the specific behaviors.

Part 2: The Characteristics of the Science Teaching

Interpret these as outlined by Fischler and Zimmer (34). The section labelled "Characteristics" require decisions ranging on a continuum. Fischler and Zimmer's interpretations of this section is given below:

Characteristics of Teaching

1. Concrete Abstract: At the abstract end of this continuum, the instructor uses verbal and written symbols, while at the other end the real object is used. In between will be various substitutes such as models, pictures, diagrams, mock-ups, etc. This continuum refers to the method of communication which the teacher used to impart knowledge and understanding. In practice, the abstract end will be checked only if the teacher and/or students talk and/or write; the concrete end, only if the real object is used. Use of any of the various A-V aids will place the check in the middle.

2. Practical Theoretical: This is the continuum which, at one extreme, makes use of applications or principles to subject matter which is within the experience of the learner and which is used by him; at the other extreme, theory is presented with no application to the learner's experience. This continuum has to do with the subject matter taught. If there is little or no mention of theory or scientific principles, the lesson will be considered practical; if no mention is made of practical applications, it will be theoretical.

3. Directed Non-directed: This continuum applies only to student activities. By this is the meant the extent to which the teacher actively directs the students in their actions. At one extreme, the teacher gives very explicit directions, frequently in writing; at the other, the student is left on his own. (34)

Recording

RECORD DOMINANT PROCESS IN PART 1 AND DOMINANT TECHNIQUE IN PART 2 AS FOLLOWS:

If activity takes place during the first part of the time interval ⧅

If activity takes place during the second half of the time interval ⧄

If activity takes place during the whole of the time interval ☒

Record activities observed which are incidental to the dominant one as ☑

APPENDIX D

Sample Pages From Experimental Curriculum

T-A-1

TEACHER'S NOTES

Introduction to Life Science Course

The first week or so should be devoted to an introduction to the course. The following points should be covered in this introduction:

1. An overview of science and scientific methods.
2. A brief explanation of the inquiry method.
3. An outline of procedures to be used in organizing groups, setting up stations, using laboratory manuals, etc.
4. An outline of the course to be covered in life science.
5. A brief survey of the processes in scientific inquiry. It may be worthwhile to prepare booklets which explain each of the processes and pass these out to the students so that they are readily available.
6. Arrangements should be made for the class to bring in collections of plant and animal life in their communities. Some attention should be given to the techniques of collecting, mounting, preserving, etc. of specimens.

DIFFERENT COMMUNITIES EXIST IN THE BIOSPHERE

INVESTIGATION #1

Concept - There is wide variety of life, or many organisms are distributed and occupy many environments. Plants and animals are adapted and distributed in relation to geographic and environmental factors.

Subconcept - There are many different communities.

Processes Involved

In this first investigation only four processes are being used. The teacher should relate the processes to the content of the investigation to as great an extent as possible. Students should be encouraged to consult the booklet which explains the processes whenever necessary.

I. Preparation

1. Problem

Teachers may wish to draw an analogy between the macro and micro communities which exist among the human population of the world and the similar types which exist among the plant and animal populations of the world.

2. Background Information

Students should be provided with an opportunity to do literature research in the classroom library and in the school library. Since this is the first occasion on which the students are required to search through a variety of reference books, an excellent opportunity is provided for the teacher to give attention to reading skills such as skimming, getting the main idea, summarizing, organizing, etc. and to library skills.

In addition to literature research, the background information could be developed through such aids as films and filmstrips. Some suggested reference texts and audio-visual aids are listed below:

Reference Texts:	World of Living Things (Brandwein) - P. 192-242
	Life: Its Forms and Changes (Brandwein) - P. 4-68
	Life Science - A Modern Course (Mason and Peters) - P. 306-327
	Living Things (Fitzpatrick, Bain) - P. 91-142

Reference Texts: (continued) Exploring Life Science (Thurber & Kilburn) - P. 9-59.

A - Different Communities Exist in the Biosphere

Investigation A - 1

Problem:

What are the different communities that exist in the biosphere?

Background Information:

In your research attempt to determine what

a) a biosphere is

b) a community is

Having determined the meanings of these two terms attempt to identify the various communities within the biosphere, their characteristics, and the plants and animals within each of these communities.

Interpretation of Data:

What is a biosphere?

What is a community?

Reference Texts:
High School Biology (BSCS-Green Version) - p. 2-10
A Sourcebook for the Biological Sciences (Brandwein)
Modern Life Science (Fitzpatrick, Hole) - p. 130
Nuffield Biology - Text III and IV

Filmstrips: (Available from Audio-Visual Aids Branch of Department of Education.)

2168 - How Animals Live in the Desert
2169 - How Animals Live in the Arctic
2170 - How Animals Live in the Grasslands
2171 - How Animals Live in the Sea
3542 - The Field as a Community
3544 - The Forest as a Community
3545 - The Pond as a Community
3750 - Nature's Half Acre
4358 - The Grasslands - A Major Community
4359 - The Swamp
4360 - The Desert
4361 - The Seashore
4362 - The Pond
4363 - The Forest
2580 - The Woods of Home

Films: (Available from IMC)

D-F68-4 Common Animals of the Woods
The Community

IV. Conceptualization of Data

11. Interpreting Data

The intention here is to have the students summarize the different communities, their characteristics and the plant and animal life found within each community. As a result of their research students should realize that there are possibly three basic macro communities: fresh water communities, marine communities and land communities. Within each of these larger communities are numerous smaller communities: pond communities, lake communities, stream communities, shallow sea communities, deep-sea communities, desert communities, forest communities, grassland communities, tundra communities, etc. Films and/or filmstrips may be used to review the more common communities or investigate the more unusual communities.

V. Open-endedness

It is hoped that the teacher will provide for individual differences and enrichment for the more capable students through activities such as seeking further evidence, identifying new problems and applying discovered knowledge.

APPENDIX E

The Attributes of a Good Teacher

A DESCRIPTION OF THE TEACHERS BASED ON INTUITIVE JUDGMENT: A RANKING ON DIFFERENT ATTRIBUTES

CONFORMING * FUNCTION	EQUAL	NONCONFORMING FUNCTION
	1. Explaining, informing showing how	
D	A, C, and E	B
	2. Initiating, directing, administering	
D	A, C, and E	B
	3. Unifying the group	
D A	C E	B
	4. Giving security	
D C	A E	B
	5. Clarifying attitudes, beliefs, problems	
D A	C and E	B
	6. Diagnosing learning problems	
D	A, C, and E	B
	7. Making of curriculum materials	
	B, C, D, and E	A

---

*Conforming to the attributes of a good teacher as described by Powley (94).

CONFORMING FUNCTION	EQUAL	NONCONFORMING FUNCTION
	8. Evaluating, recording, reporting	
D	A, B, C, and E	
	9. Organizing and arranging classroom	
C D	E B	A
	10. Rapport and interaction	
D A	C B	E
	11. Student respect for teacher	
	A, C, D, and E	B
	12. Teacher respect for students	
D	A and C E	B
	13. Child open to his experience-individuality	
D	A and E B	C
	14. Communicative Skills	
D	A and C E	B
	15. Responsive to the data the child and group are placing in the situation	
D A	B, C, and E	

CONFORMING FUNCTION	EQUAL	NONCONFORMING FUNCTION
	16. Exercise of controlling functions	
	A, C, D, and E	B
	17. Supportive interpersonal relationships	
D	A, B, D, and E	
	18. The processes of science and the scientific enterprise	
D	A, B, and E	C